NATURAL AGENCY: AN ECOLOGICAL APPROACH

by

Fermín Cristóbal Fulda de la Garza

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Institute for the History and Philosophy of Science and Technology
University of Toronto

© Copyright by Fermín Cristóbal Fulda de la Garza 2016
Abstract

Agency, the capacity to act for a goal or purpose guided by norms, is central to our understanding of the capacities and activities of organisms including human beings. However, its distinctive purposive and normative character has proven difficult to integrate with the scientific understanding of organisms as natural physical entities. The challenge is to show both that agency has a place in the natural causal order of the world as described by natural science (naturalism), and that its distinctive purposive and normative character plays an indispensable role in our understanding of natural phenomena (teleology). The standard approaches, however, either locate agency in the natural causal order of the world but fail to vindicate its distinctive purposive and normative character, or vice versa. My goal is to avoid this dichotomy and instead offer a complete unifying account of natural agency. In the first part of the thesis I address the methodology of naturalism. First, I argue that the standard approaches are committed to an unnecessarily stringent set of assumptions about what naturalism requires. And second, I propose an alternative methodological strategy based on scientific work in complex systems dynamics. In the second part I use this strategy to articulate a novel account of natural agency, The Ecological Theory of Agency. This account is based on recent empirical and philosophical work in evolutionary developmental biology on the organism-environment relation (Walsh 2013). An agent, I propose, is an ecologically embedded purposive system. That is, a system with the gross
behavioral capacity to bias its repertoire in response to what its conditions afford for attaining its goals. Then I show how the solution that this account of agency provides generalizes to two other instances of the dichotomy. The first is the problem of bacterial cognition. I show how the ecological approach allows us to navigate a middle way between thinking of unicellular organisms as either mere automata or full-blown cognitive agents. The second is the problem the role that reasons play in the explanation of action. I show that the ecological approach avoids thinking of reasons as either exclusively causes or exclusively norms.
Acknowledgments

First, I would like to express my deepest gratitude to my advisor and mentor Prof. Denis Walsh for his constant supervision, encouragement, and commitment to my intellectual and personal growth throughout my PhD. His generous and patient guidance has made the better parts of this thesis possible and has set a very high standard for my future research.

I would also like to thank Prof. William Seager for his support throughout the whole process and for his insightful comments along the way. In particular, I will like to thank him for his prompt response and encouragement in the final stages of my research.

I would also like to thank my fellow students Michael Cournoyea, Alex Djedovic, Eugene Earnshaw-Whyte, Cory Lewis, and Parisa Moosavi for their feedback and for providing such a friendly and intellectually enriching group.

I am also grateful to Rasmus Winther, Fernando Flores-Mangas, Nicolas Sanchez-Guerrero, Hakob Barseghyan, Andres Fulda, Oscar Espinosa and Luis Rodrigo Basaguren for many thoughtful conversations.

I owe my vocation to the early Socratic approach of my parents, Irma de la Garza García and Fermín Fulda Fernandez. They have supported and encouraged me to pursue my dreams with love and dedication. This thesis is dedicated to them, with love and admiration.

For deciding to come to Canada to share her life with me; for staying by my side through difficult and long winters; and for bearing with me throughout the PhD, I thank my dear wife Ana Paula Espinosa. Without her love, sacrifice and commitment this thesis would not have been possible.

Finally, I dedicate this thesis to my son Fermín Fulda Espinosa who has enriched and inspired me beyond measure.
# Table of Contents

Acknowledgments ........................................................................................................ iv

Introduction .................................................................................................................. x

Part 1 The Methodology of Naturalism ......................................................................... xiii

Chapter 1 The Problem of Natural Teleology ................................................................. 1

1.1. Introduction .............................................................................................................. 1
1.2. Two Desiderata ........................................................................................................ 3
   1.2.1. Explanatory Adequacy .................................................................................. 3
   1.2.2. Naturalism .................................................................................................... 6
1.3. The Problems of Natural Teleology ....................................................................... 7
   1.3.1. The Argument from Non-Actuality ............................................................... 7
   1.3.2. The Argument from Intentionality ............................................................... 8
   1.3.3. The Argument from Normativity ................................................................. 8
   1.3.4. The Argument from Explanatory Exclusion ............................................... 8
1.4. The Standard Approaches ..................................................................................... 9
   1.4.1. Etiology ......................................................................................................... 9
   1.4.2. Mechanism ..................................................................................................... 14
      1.4.2.1. Cybernetics ............................................................................................ 14
      1.4.2.2. Systems-Theory ..................................................................................... 16
      1.4.2.3. Autonomous Systems ......................................................................... 17
   1.4.3. Normativism .................................................................................................. 19
      1.4.3.1. Axiology ................................................................................................. 20
      1.4.3.2. Rationality ............................................................................................. 22
1.5. Conclusion: A Dichotomy between Eliminativism and Primitivism ....................... 24
Chapter 2 Two Dogmas of Naturalism ................................................................. 26
  2.1. Introduction ............................................................................................... 26
  2.2. The First Dogma of Naturalism: Individuation is causal realization ........ 28
      2.2.1. The Causal Theory of Perception ..................................................... 31
      2.2.2. The Causal Theory of Mental Content .......................................... 32
      2.2.3. The Causal Theory of Knowledge .................................................. 33
      2.2.4. Summary and Evaluation ............................................................... 34
  2.3. The Second Dogma of Naturalism: Explanation is Causal Realization ..... 34
      2.3.1. The Molecular Gene ...................................................................... 35
      2.3.2. The Causal Theory of Action .......................................................... 36
      2.3.3. Summary and Evaluation ............................................................... 38
  2.4. Eliminativism and the Two Dogmas ....................................................... 38
      2.4.1. Eliminativism and the First Dogma ............................................... 39
      2.4.2. Eliminativism and the Second Dogma .......................................... 39
      2.4.3. How the Two Dogmas Generate Eliminativism ............................. 40
  2.5. Primitivism and The Two Dogmas ........................................................... 42
      2.5.1. Primitivism and the First Dogma .................................................... 42
      2.5.2. Primitivism and the Second Dogma ............................................... 43
      2.5.3. How the Two Dogmas Generate Primitivism ................................. 44
  2.6. Conclusion ................................................................................................ 45

Chapter 3 Three Grades of Naturalistic Involvement ..................................... 47
  3.1. Introduction ............................................................................................... 47
  3.2. Grade 1: Causal Realization .................................................................... 48
      3.2.1. Transubstantiation ....................................................................... 48
      3.2.2. Demonic Possession .................................................................... 49
3.3. Grade 2: Scientific Role .......................................................... 50
   3.3.1. Thermodynamics and Statistical Mechanics .................. 50
   3.3.2. Fluid Dynamics ............................................................. 54
   3.3.3. Population Genetics ...................................................... 57
   3.3.4. Lessons from G2: Naturalism and Emergence .................. 59
3.4. Grade 3: Scientific Essentialism ........................................ 61
   3.4.1. Dispositional Essentialism ............................................ 62
   3.4.2. Chemical Microstructural Essentialism ......................... 63
   3.4.3. Developmental Microstructural Essentialism .................. 64
3.5. Conclusion: Naturalism without Dogmas .............................. 66

Part II Agency as an Ecological Concept .................................. 68

Chapter 4 An Ecological Theory of Natural Agency ..................... 69
   4.1. Introduction ................................................................. 69
   4.2. The Phenomenon of Agency ............................................ 70
      4.2.1. The Squirrel and the Acorns .................................... 70
      4.2.2. The Mechanical and the Ecological ................................ 70
   4.3. The Theoretical Indispensability of the Ecological Conception ... 73
      4.3.1. Empirical Basis ....................................................... 73
      4.3.2. Predictive Role ...................................................... 76
      4.3.3. Explanatory Role .................................................... 77
      4.3.4. Normative Role ...................................................... 77
   4.4. G2 Naturalization of Goals ........................................... 78
      4.4.1. The Causal Realization of Goals ................................. 79
      4.4.2. The Role of Goals in the Individuation of Agents ............ 79
         4.4.2.1. Situated Adaptationism ....................................... 79
Chapter 4

4.4.2.2. Goals .................................................................................................................. 81
4.4.2.3. Repertoire ......................................................................................................... 82
4.4.2.4. Affordances ...................................................................................................... 83
4.4.3. The Role of Goals in Teleological Explanation ....................................................... 88
4.5. Conclusion ................................................................................................................ 91

Chapter 5 The Case of Bacterial Cognition ..................................................................... 94

5.1. Introduction ............................................................................................................... 94
5.2. The Cognitive Turn in Microbiology and Cell Biology ............................................ 95
   5.2.1. Bacteria are Purposive Agents ........................................................................... 95
   5.2.2. Purposive Agency is the Mark of Cognition ...................................................... 98
   5.2.3. The Role of Cognition in Microbiology .......................................................... 102
5.3. Cartesian Agency ..................................................................................................... 104
   5.3.1. The Mechanistic Turn and the Primacy of Cognition ...................................... 104
   5.3.2. The Dilemma between Mechanism and Intellectualism ................................. 106
   5.3.3. Minimal Cognition .......................................................................................... 109
5.4. Ecological Agency ................................................................................................... 111
   5.4.1. Agency as an Ecological Concept ..................................................................... 111
   5.4.2. A Continuum of Natural Agents ....................................................................... 112
5.5. Conclusion ............................................................................................................... 116

Chapter 6 The Case of Action Explanation ..................................................................... 118

6.1. Introduction ............................................................................................................... 118
6.2. The Problem of Action Explanation ........................................................................ 119
6.3. The Standard Accounts and Their Discontents ........................................................ 120
   6.3.1. The Causal Theory of Action ......................................................................... 120
   6.3.2. The (Anti-Causal) Teleological Theory of Action ........................................... 123
6.3.3. A Dilemma between Eliminativism and Primitivism (again) ......................126

6.4. Diagnosis ........................................................................................................127

6.4.1. Why CTA leads to Eliminativism .................................................................127

6.4.2. Why TTA leads to Primitivism ...................................................................129

6.4.3. A Model-Based Compatibilism .................................................................131

6.5. Towards an Ecological Theory of Action .......................................................132

6.6. Conclusion ........................................................................................................137

Bibliography .............................................................................................................138
Introduction

My dissertation offers an ecological account of natural agency and teleological explanation. This account is based on recent empirical and philosophical work in complex systems dynamics and ecological psychology, particularly as applied to developmental biology. I argue that teleology plays an indispensable role in the explanation of the gross ecological dynamics of naturally occurring agents. Agents are purposive systems characterized by their capacity to respond to conditions as affordances, that is, as means to attain their goals.

The dissertation has two parts. The first part includes the first three chapters. It is about the methodology of naturalism. The second part includes the last three chapters. It applies the new methodology to the problem of naturalizing agency and teleology.

In the first chapter I introduce the problem of saying what goals are and how they can explain anything. I propose two desiderata, naturalism (physicalism) and explanatory adequacy (purposes must explain as purposes). I argue that while standard etiological and mechanistic approaches satisfy naturalism at the expense of explanatory adequacy, normativist approaches do the opposite. This dialectic, I conclude, leads into a dichotomy between eliminativism and primitivism.

In the second chapter I offer a diagnosis of this dichotomy. I argue that the standard approaches are predicated on two stringent assumptions about what the method of naturalization requires. I call them ‘two dogmas of naturalism’. The first dogma (D1) is the view that to say what some phenomenon is just is to say how it is causally realized. And the second dogma (D2) is the view that the theoretical role of a phenomenon can be specified exhaustively by citing its causal realizer. The dichotomy is generated, I conclude, by the implicit commitment of eliminativism to these dogmas and by their rejection by primitivism.

In the third chapter I distinguish ‘three grades of naturalistic involvement’, each of which subsumes the previous one and provides increasingly stringent criteria for a phenomenon to be natural. According to the first grade (G1) to be natural is to be realized in the causal order of the world. In turn, for G2 to be natural is to be causally realized and to play a distinctive theoretical role in science irrespective of its causal realizer. Finally, for G3 to be natural is to be causally
realized, play a distinctive theoretical role is science and for that role to be specified in terms of its causal realizer. G2 and G3 are both legitimate ways of giving a scientific naturalization. But only G3 is committed to the two dogmas. I conclude that the standard G3 strategy is not mandatory. The remainder of the dissertation offers a G2 naturalization of agency and teleology.

In the fourth chapter I apply G2 to the phenomenon of agency and teleology. Agency, I propose, is an empirical phenomenon. An agent is a purposive system with the gross behavioral capacity or ‘repertoire’ to respond appropriately to its network of ‘affordances’. An affordance is a condition of existence that prompts an appropriate response and hence is relevant or significant for the attainment or maintenance of a ‘goal’. I call this the ecological theory of agency. There is no G3 analysis of ‘purpose’, ‘repertoire’ and ‘affordance’; they make a tight, interdefined cluster of concepts indispensable for understanding an agent’s behavioral dynamics. As for teleology, I appeal to a recent ‘invariance’ account according to which the role that goals play in explanation can be specified in terms of a distinctive relation of counterfactual dependence (Walsh 2012).

The last two chapters apply the ecological account to instances of the eliminativist vs. primitivism dichotomy in the philosophy of microbiology and the philosophy of action respectively.

In the fifth chapter I contrast the ecological approach to the traditional Cartesian conception using recent research in microbiology on bacterial cognition and cell-decision-making as a test case. I argue that the Cartesian conception forces us to choose between attributing full-blown intentional psychology to bacteria and treating them as mere machines. Most organisms, however, occupy a middle ground between these two extremes. In contrast, since the repertoire of a purposive system can vary in its richness, the ecological approach allows for degrees of natural agency along a continuum, from mere adaptive biological agents at one extreme to cognitive psychological agents on the other. This view makes it possible to conceive unicellular organisms as minimal biological agents, hence not as mere machines, but without committing us to the view that they are cognitive (rational) agents.

The last chapter recapitulates the whole thesis. I propose an ecological account of reasons for action as a way to overcome the debate between causalists and (anti-causal) teleologists about the role of reasons in the explanation of action. I argue that the dialectic between causalism and the teleologism leads into a familiar dilemma between eliminativism and primitivism about the
normativity of action explanation. As an alternative I argue that the theoretical role of reasons in the explanation of an action is to give the *ecological role* of the action. The ecological role of an action, in turn, is the contribution it makes to the attainment of agent’s goal given her repertoire and affordances. This approach vindicates both the fact that actions are done and hence caused by agents, and the fact that they are done for a reason that justifies the action. On account of this, I conclude, the ecological approach avoids the dilemma and offers a complete and naturalistic account of the causal and normative role that reasons play in the explanation of action.
Part 1
The Methodology of Naturalism
Chapter 1
The Problem of Natural Teleology

1.1. Introduction

In this chapter I introduce the problem of natural teleology, the standard approaches to naturalizing teleology and pose a dichotomy for these approaches. I take teleology to be a particular mode of explanation in which the occurrence of some event or property is explained by citing the goal, purpose or end that it subserves (Walsh 2008). The standard approaches tend to conflate the project of naturalizing teleology with the project of naturalizing agency. But even if I take an account of natural teleology to be part of an account of natural agency (see Chapter 4), I believe these projects should be kept distinct. Teleology may be the mark of agency. After all, teleology is the only way to explain an agent’s actions as actions. But teleology is not sufficient for agency. Some purposive systems, such as the immune or the thermoregulatory system are susceptible to teleological explanation but they are not agents. Similarly, we can explain the behavior of artefacts teleologically but they are hardly agents. Furthermore, I have methodological reasons for observing the distinction between teleology and agency. The approach I will take to teleological explanation in Chapter 4 exemplifies a distinctive strategy of naturalization. This strategy generalizes to agency. But the existent approaches to teleology exemplify other strategies. And as it will become apparent in Chapter 3, I believe these strategies fail on methodological grounds.

Teleology explains why something occurs by citing its goal. But what is a goal? And how can goals explain anything? As a mode of explanation teleology has very distinctive features from the causal-mechanical explanations commonly used in natural science. Yet teleological explanations readily apply to natural beings such as biological organisms and sub-organismic systems as well as intentional agents and the artefacts they create. An adequate account of teleological explanation must thus show both that goals have a place in the natural order of things as described by natural science, and that they play a distinctive role as goals in the explanation of natural phenomena. I call these desiderata ‘naturalism’ and ‘explanatory adequacy’ respectively.
I distinguish three main standard approaches to naturalizing teleology: The etiological approach, the mechanical approach, which includes the cybernetic, systems-theoretic, and autonomous-systems accounts, and the normative approach, which includes the axiological and the rationality accounts. I will argue that despite their merits they fail to satisfy both desiderata.

The strategy of the etiological approach is to reduce teleological explanation to functional explanation, a distinctive form of causal explanation. But despite their superficial similarities, I argue that functional explanations are not teleological explanations. So etiology naturalizes teleology by elimination. Also, as a reductive approach, the etiological approach is willing to abandon some of our pre-theoretic intuitions about teleological explanation. But my project seeks to save the phenomena and hence vindicate our pre-theoretic intuitions.

In contrast, the strategy of mechanism is to start with a naturalized account of what it is to be purposive system and then derive an account of teleological explanation on that basis. I argue that while this approach offers a good account of what it is for a natural entity to be a purposive system, it fails to vindicate the distinctive role that goals play as goals in teleological explanations. Hence, it is also an eliminativist account of natural teleology. Furthermore, this strategy assumes that an account of teleological explanation simply follows from an account of purposiveness. But I take these to be independent yet importantly related projects (as it will become apparent in Chapter 4). So mechanism will also prove inadequate for my project.

Finally, the strategy of the normative approach is to start with an account of the concept of a goal in irreducibly normative terms, and then derive an account of the distinctive explanatory role that goals play as goals on that basis. I argue that this approach is explanatorily adequate as it vindicates the distinctive features of teleological explanation. However, by taking goals to be a primitive feature of the world, this approach is silent about their place in the natural order. Hence, it fails to satisfy naturalism.

While etiology and mechanism satisfy naturalism at the expense of explanatory adequacy, normativism does the opposite. So whatever merits the existent approaches have none provides a unifying theoretical account of the kind I’m looking for. This dialectic, I conclude, leads into a dichotomy between eliminativism, the view that goals have a place in the natural order of things but they play no distinctive explanatory role as goals, and primitivism, the view that goals play a distinctive explanatory role as goals but there is no account of their place in the natural order.
In section 1.2 I introduce the two main desiderata for a theory of teleology, naturalism and explanatory adequacy. In section 1.3 I elaborate on these desiderata by introducing the standard objections against natural teleology. In section 1.4 I introduce the existent approaches and assess them in light of the desiderata. Finally, in section 1.5 I argue that the dialectic between the standard approaches leads into a dichotomy between eliminativism and primitivism.

1.2. Two Desiderata

Teleology explains why something occurs by citing its goal. But what is a goal? And how can goals explain anything? On the one hand teleology applies to natural entities such as organisms and their parts. An adequate account of natural teleology must thus show that goals have a place in the natural order of the world. Call this desideratum naturalism. On the other hand teleology has very distinctive features from those of causal-mechanical explanations used in natural science. So an adequate account must capture the distinctive role that goals play as goals in explanation. Call this desideratum explanatory adequacy.

1.2.1. Explanatory Adequacy

Teleology is a mode of explanation in which the presence, occurrence or nature of some event or property is explained by citing the goal, end or purpose to which it contributes (Walsh, 2008). For example, knowing that a person had the goal of exercising explains why she was running so early in the morning. Similarly, knowing that the goal of the immune system is to restore the health of an organism explains why it produced a certain amount of lymphocytes (white blood cells). We also explain why a given mammal sweats by citing the goal of its thermoregulatory system to maintain its temperature within a certain range. Where ‘S’ denotes a purposive system, ‘ϕ’ denotes the means, ‘in order to’ denotes the telic connective and ‘ψ’ denotes the goal, 1 below gives the form of a teleological explanation (Wilson 1989):

1. $S$ did/does $ϕ$ in order to $ψ$ (or $S$’s $ϕ$-ing was directed at $ψ$)

The idea is that 1 is true if and only if $S$ did/does $ϕ$ because doing $ϕ$ contributes to the fulfillment of $ψ$. $ϕ$, the means, is the explanandum while $ψ$, the goal, is the explanans. Teleology can take an explicit psychological form as per 2:
2. $S$ did $\phi$ because $S$ desires or intends to $\psi$ and $S$ believes or thinks that doing $\phi$ is the/a means to $\psi$.

Here the goals and the means are described in terms of the contents of the agent’s intentional mental states: the goals are represented as the content of conative states such as desires or intentions while the content of cognitive states such as instrumental beliefs and thoughts represent what the agent takes to be the means to attain her goals. But not all teleological explanations are intentional. Teleological explanations in which the purposive system is a biological sub-system, such as the immune system or the thermoregulatory system for example are not intentional.

Teleological explanations capture a distinctive relation between the means or explanandum and the goals or explanans that Aristotle called “hypothetical necessity” (Cooper 1987). Roughly, $\phi$ is hypothetically necessary for $\psi$ just in case $\psi$ is a goal and $\phi$ is necessary for [or no worse than anything else/or good enough for] the attainment of $\psi$ under the circumstances $C$ (Walsh 2008). The idea is that a purposive system $S$ would typically do what is (hypothetically) necessary to attain its goals $\psi$ across a range of conditions $C$, that is, across a range of values for $\psi$ and $C$. The idea is that something is hypothetically necessary if some goal is to be attained. For example, according to Aristotle (see Cooper 244) an axe has to be hard in order to perform its function of cutting or splitting. And in order to be hard the axe has to be made of the appropriate material, such as iron or bronze. So it is (hypothetically) necessary for a thing to be made of iron or bronze if it is to be an axe. Similarly Aristotle holds that human beings must have eyes made of fluid material and covered with flaps of flesh (eyelids). Being made of the relevant material is thus hypothetically necessary to achieve the goal of sight. So anything that is necessitated hypothetically is necessitated as a means to a goal.

The relation of hypothetical necessity between goals and means can form a hierarchy in which a goal at a lower level in the hierarchy is a means for another goal at a higher level in the hierarchy. For example, organisms may be said to have the ultimate goal of surviving and reproducing. To attain these ultimate goals they have many intermediate goals such as eating, mating, sleeping, etc. Finding a prey is a goal, but is also a means to attain the goal of eating, which in turn is a means to attain the goal of surviving which in turn is a means to attain the goal of reproducing.
Means cause their ends, but the relation between means and ends is not a mere relation of cause to effect, it has a different “modal profile” from the relation of causation. The modal profile of hypothetical necessity can be characterized in terms of the counterfactuals 3 and 4 that it supports (Sehon 1994; Walsh 2012):

3. If the fulfillment of $\psi$ had required $\Theta$-ing (rather than $\phi$ –ing) then (ceteris paribus) $S$ would have $\Theta$-d.

4. If the goal of $S$ had been $\psi^*$ rather than $\psi$ then (ceteris paribus) $S$ would have $\phi^*$–d.

For example, if it had been raining in the morning then (ceteris paribus) instead of running the person would have exercised in some other way, just as if the mammal had been cold rather than warm the thermoregulatory system would have triggered the shivering reflex to maintain homeostasis rather than the sweating reflex. Similarly, if her goal had been to rest rather than to exercise she wouldn’t have been running in the morning, she would have stayed at home sleeping. These counterfactuals capture the characteristic flexibility and robustness of purposive systems, namely, their capacity to attain their goals reliably across a wide range of conditions by producing appropriate responses in the face of perturbations.

Hypothetical necessity has distinctive normative implications. Goals require their means (Walsh 2012) and means are supposed to be a good for or appropriate for their goals. The goal of a system imposes on it the requirement to bring about the means to its attainment given the circumstances. So 1 implies 5 below:

5. If $S$ has goal $\psi$ and $\phi$ –ing is the best means to attain $\psi$ or is good for attaining $\psi$ then (ceteris paribus) $S$ should $\phi$.

On this basis we can evaluate whether $\phi$ –ing is appropriate or inappropriate, that is, good for or not good for attaining $\psi$ under the relevant circumstances $C$. A teleological explanation thus implies, ceteris paribus, the following evaluative judgment:

6. Doing $\phi$ was appropriate/good for attaining $\psi$ in $C$.

For example, given her goal of exercising and the circumstances she ought to or is required to bring about the means to exercise, just as keeping an organism’s health imposes a requirement on
the immune system to produce the right amount of lymphocytes. This normativity expresses the fact that purposive systems are characterized by their capacity to respond appropriately to their circumstances guided by norms or requirements that their goals impose on them. An explanatorily adequate account of teleological explanation must thus vindicate the distinctive normative and modal profile of hypothetical necessity.

1.2.2. Naturalism

Given the distinctive modal and normative features of teleological explanation, the challenge for a naturalized account of teleological explanation is to show that teleology can be brought under the ambit of naturalism. Naturalism is the view that nature, understood as the causally complete immanent order of the world as described by natural science, exhausts reality such that natural science provides the best method for understanding the nature of reality (Papineau 2009; Quine 1969). Sellars (1963, 173) offers a bold statement of the priority of science characteristic of naturalism:

“In the dimension of describing and explaining the world, science is the measure of all things, of what is that it is, and of what is not that it is not”.

Insofar as philosophy deals with the nature of reality and our epistemic access to it naturalism implies that philosophy must be continuous with natural science. As Quine (1981, 21) puts it, naturalism is

“the recognition that it is within science itself, and not in some prior philosophy, that reality is to be identified and described”.

We can distinguish two kinds of naturalism, metaphysical and methodological. Methodological and metaphysical naturalism can be taken to be logically independent. I will take naturalism to include both. Methodological naturalism is a view about how to investigate reality according to which the scientific method offers the best guide (Chomsky 2000; Giere 2006). Metaphysical naturalism, on the other hand, is a view about the content of reality. As Dewey (1939, 580) puts it:
“Naturalism is opposed to idealistic spiritualism, but it is also opposed to super-naturalism and to that mitigated version of the latter that appeals to transcendent a priori principles places in a realm above Nature and beyond experience.”

Metaphysical naturalism is committed to physicalism, the view that the physical facts and properties exhaust reality such that every fact or property is either identical to or realized by physical facts and properties (Kim 1993; 2005; Shoemaker 2007). The idea is that \( x \) realizes \( y \) just in case the existence of \( x \) is constitutively sufficient for the existence of \( y \) such that \( y \)’s existence is “nothing over and above \( x \)’s existence” (Shoemaker 2007, 4, 6). Physicalism is motivated by the principle of the causal closure or the causal completeness of the physical world, the view that “no physical event has a cause outside the physical domain” (Kim 1993, 280). This principle implies that anything that makes a causal difference in the physical world must itself be physical or physically constituted. Thus, the physical world is self-contained.

1.3. The Problems of Natural Teleology

At least since the inception of the scientific revolution in the XVII century the standard view has been that teleology is incompatible with naturalism. This implies that the desiderata of naturalism and explanatory adequacy constitute an incoherent set of requirements. Here I review four prima facie objections that any naturalistic account of teleology must meet (Walsh 2008).

1.3.1. The Argument from Non-Actuality

A teleological explanation explains the occurrence of some event by showing it to be the (or a) means to the attainment of some goal. Citing the goal as a goal explains why—as opposed to how—the means occurred or tends to occur. Typically, however, at the time of the occurrence of the means the goal is an unactualized, future state of affairs. In fact, it is the occurrence of the means that causes the fulfillment of the goal. But how can an unactualized future state of affairs explain anything? The worry is that teleological explanations are committed to some mysterious form of causation by unactualized future state of affairs, such as backwards causation. This commitment is of course incompatible with naturalism. An adequate account of natural teleology must thus show that teleology is not committed to causation by non-actualia.
1.3.2. The Argument from Intentionality

To avoid commitment to causation by non-actualia we can appeal to intentionality, the capacity of the mind to represent the world. The idea is that it is not the unactualized future state of affairs \textit{per se} that explains the occurrence of the means. Rather, it is the internal representation of the unactualized future state of affairs that (efficiently) causes and hence explains the occurrence of the means. After all, as the paradigmatic case of teleology in psychological contexts indicates, we can explain the occurrence of the means by appealing to the intentions of the agent. But although this avoids the problem of causation by non-actualia, it severely limits the scope of teleology, as most organisms are not intentional agents and neither are sub-organismic parts. An adequate account of natural teleology must thus be able to avoid reducing teleology to intentionality.

1.3.3. The Argument from Normativity

As we have seen, explanations that cite goals have normative implications. They tell us what the agent \textit{ought} to do given her goals and circumstances. But for the means to be something that the agent ought to produce, the goal, some argue, must be a state of affairs that ought to be attained. But a goal could not be a state of affairs that ought to be attained unless the goal was \textit{good} (Bedau 1992). After all, it is the value that a given state of affairs has for the agent that imposes an obligation or demand to bring that state of affairs about. The worry is then that teleology is committed to the existence of intrinsically evaluable state of affairs, which is incompatible with naturalism. So an adequate account should accommodate teleology’s normative implications without positing intrinsically evaluable state of affairs.

1.3.4. The Argument from Explanatory Exclusion

According to the principle of explanatory exclusion (Kim 1989), no event can have more than one complete an independent (causal) explanation. Together with the principal of causal-closure we get the conclusion that teleology is dispensable: (1) Every event has a complete mechanistic cause; (2) every event has a complete mechanistic explanation; (3) every event has at most one complete explanation; (4) mechanistic explanation is complete and exhaustive; (5) for there to be a role for goals in science they would have to cause phenomena that mechanism cannot. Hence,
there is no role for goals in scientific explanation. An adequate account of teleology must thus show both that goals play an indispensable role as goals, and that this role is compatible with the completeness of mechanism.

In summary, a naturalistic account of teleological explanation must:

(i) Avoid commitment to causation by non-actualia.

(ii) Avoid reducing teleology to intentionality.

(iii) Capture the distinctive normative features of teleological explanation without positing intrinsically evaluative states of affairs.

(iv) Show how teleology can explain something that mechanism can’t given the explanatory completeness of mechanism.

1.4. The Standard Approaches

I will now introduce the standard approaches to the problem of naturalizing teleology and test them against the desiderata of naturalism as per (i)-(iv) and explanatory adequacy as per the normative and modal profile of hypothetical necessity. The standard approaches fall into three main categories. I call them the ‘etiological’, the ‘mechanistic’ approach and the ‘normative’ approach. Etiology is an explicitly eliminativist approach to teleological explanation. It holds that teleological explanations are a special kind of causal explanation. Mechanism, in turn, is a set of attempts to account for the nature of purposiveness. The motivating idea is that an account of purposiveness can be used as the basis of an account of teleological explanation. Finally, the normative approach is a set of attempts to account for the nature of teleological explanation in irreducible normative terms.

1.4.1. Etiology

The etiological account was constructed precisely to naturalize the distinctive explanatory character of teleology. The strategy is to reduce teleological explanation to functional explanation, a distinctive form of causal explanation. The function of an entity is the effect that explains the presence of that entity (Wright 1973): x has the function y if and only if:
To say that ‘x is for y’ is to say that ‘x is there because it does y’. Functional attributions offer a distinctive form of causal explanation in which the presence of something is explained by citing its effects. They support a distinction between functional effects that explain the presence of x and accidental effects that don’t. They also support a distinction between function and malfunction: ‘x is for y’ does not entail ‘x does y’ so functional attributions are non-factive.

When ‘x does y’ is false x is said to malfunction. This distinction is supposed to be normative: ‘x is for y’ entails that ‘x is supposed to y’ such that, if ‘x does y’ is false the x is not doing what it is supposed to do.

This form of explanation applies to artefacts as well as actions. For example, we say that the fuel warning light is there because it indicates when the fuel is empty (1); and indicating when the fuel is empty is a consequence of the light being there (2). The light might also illuminate the board or flash in synchrony with the music. But these are mere accidental effects with no explanatory import. Also, even if this fuel warning light does not indicate that the fuel is empty it should for it still has the function to do so.

Functional explanation also applies to biological traits that are adaptations evolved by natural selection. For example, we say that the heart is there because it pumps blood (1); and pumping blood is a consequence of the heart being there (2). The evolutionary function of the heart is to pump blood. The heart might also make pulse sounds, but these are mere accidental effects with no evolutionary value and hence with not causal hence explanatory import. Also, if a particular heart does not pump blood on a given occasion then it’s failing to do what it is supposed to do given what it was selected for in evolutionary history.

This account was originally intended as an analysis of the logical structure of teleological explanations, not as a naturalization of teleology (Wright 1973). Thus, while artefacts meet conditions 1 and 2 as a consequence of the intentions of a designer, biological traits meet them as a consequence of the historical process of evolution by natural selection. However, on the assumption that functional explanations are clear instances of teleological explanation, the evolutionary account of functional explanation has been taken to give a naturalization of
teleology. The basic idea is that the function or ‘purpose’ of a trait is the effect for which it was selected in the past (Millikan 1984; 1989; Neander 1991; Ruse 1971). The central notion here is that of ‘selection for’ (Sober 1984, 100): Trait type $X$ in population $P$ was selected for doing $Y$ in historical context $C$ is true if and only if past tokens of $X$ causally contributed to the fitness of individuals who possessed $X$ in $C$ by doing $Y$ and those individuals who didn’t have $X$ were on average less fit than those individuals that did. In contrast, to say that there was mere selection of trait $X$ in $P$ is to say that individuals with $X$ have a higher average fitness than individuals who lack $X$.

The etiological account thus claims to offer a non-circular strictly causal translation-schema for teleological explanation and hence to show that teleology is really just an instance of causal functional explanation. So for example, to say that ‘an organism’s immune system produced a certain amount of lymphocytes in order to restore its health’ or that ‘an organism sweated because sweating is a means to maintain its temperature stable’, which are genuine instances of teleology, can be translated without loss as ‘the immune system is an adaptation for which producing lymphocytes is useful’ and ‘the thermoregulatory system is an adaptation for which sweating is useful’ respectively, which are evolutionary functional explanations.

This translation has the virtue of eliminating the problematic future reference of teleological explanations to unactualized state of affairs that has historically motivated the objection of backwards causation in favour of a past reference. So although the historical version refers explicitly to a future effect of a trait, this future reference is just an implicit reference to the causally efficacious process of natural selection (Neander 1991). Evolutionary functions are past selected effects that cause present effects of the same type. This is an historical instance of efficient causation. So the etiological account indeed avoids commitment to causation by non-actualia and hence satisfies (i). And since the evolutionary function of a trait is specified in terms of the process of evolution by natural selection rather than in terms of the intentions of a designer, the etiological account avoids reducing teleology to intentionality and hence satisfies (ii).

Evolutionary functions are supposed to have normative implications. A heart that fails to pump blood is malfunctioning and hence is not doing what it is supposed to do given its causal-selective history. This normative implication is grounded in the process of evolution by natural
selection, a non-normative strictly causal process. So the etiological account avoids positing intrinsically evaluative properties to ground the normativity of evolutionary functions.

However, the alleged normativity of functions does not capture the characteristic normativity of teleological explanations. The normativity of teleological explanations is grounded in the relation of hypothetical necessity between goals and means. The idea is that goals require their means and means are appropriate for or good for their goals given the circumstances. So the fact that S has goal G and doing x is the means to attain G under circumstances C imposes a normative requirement on S to bring x about. On this basis we can evaluate whether x was appropriate or good for attaining G in C. Neither the relation of normative requirement between goals and means nor the relation of being good for between means and goals is an instance of the function/malfunction distinction. The function/malfunction distinction may or may not be normative. But it is not the normativity characteristic of teleological explanation. So the etiological account fails to satisfy (iii).

Functional explanations are (historical) causal-mechanical explanations. According to the etiological account teleological explanations are functional explanations. On account of this the etiological account fails to show how teleology can explain something that mechanism can’t. So this account fails to satisfy (iv). A defender of the etiological account might respond that condition (iv) begs the question: The etiological account is a reductive project and hence it is explicitly aimed at showing that teleology is not a distinctive form of explanation. However, we can turn the tables on this response and argue that it is the very project of reducing teleology that begs the question against its distinctiveness as a form of explanation. Motivated by the worry of backwards causation, the etiological account assumes that if goals are to play a distinctive explanatory role they must play a distinctive causal (efficient) role. But goals don’t explain by causing their means. Rather, it is the means that cause their goals. Goals explain by requiring their means, that is, by rendering them necessary given the circumstances for realizing the goal. So at best the reductionist aspirations of the etiological account are predicated on a reinterpretation of what teleological explanation requires and at worst it simply mischaracterizes teleological explanation.

Indeed, despite their superficial similarities, teleological explanations and functional explanations are importantly different (Walsh 2014a). First, they have different logical
structures. As we saw in section 1.2.1 teleological explanations have the following form ‘S did/does $\phi$ in order to $\psi$ (or A’s $\phi$-ing was directed at $\psi$)’. In contrast, functional explanations have the form ‘$x$ is there because $x$ does $y$; and $y$ is a consequence of $x$’s being there’. They also have different contents. Teleological explanations are about tokens, but functional explanations are about types. Evolutionary functions explain the persistence of trait types in a population by describing some effect that has a significant contribution on the fitness of that type. But a teleological explanation explains the presence of a trait token in an individual organism by describing the way it is conducive to the attainment of the biological goals of survival and reproduction.

On account of their different contents these explanations support different counterfactuals. To say that ‘an organism’s immune system produced a certain amount of lymphocytes in order to restore its health’ implies (ceteris paribus) that if restoring the health of the organisms had required a larger amount of lymphocytes the immune system would have produced them. But to say that ‘the immune system is an adaptation for which producing lymphocytes is useful’ implies (ceteris paribus) that if producing lymphocytes hadn’t been useful this trait would not have been selected (for that effect). The former captures the reliability of hypothetical necessity while the latter captures the reliability of causal necessity. So the etiological account fails to capture the modal profile of teleological explanation.

In conclusion, by identifying goals with selected effects, the etiological approach locates goals (functions) in the causal structure of the world as per naturalism. However, it fails to satisfy conditions (iii) and (iv) for naturalized teleology. And as a functional translation of teleology the etiological account assigns goals a role in explanation as causes not as goals. Hence, it cannot capture the distinctive modal and normative profile of hypothetical necessity and hence eliminates teleology in favour of causal explanation. So etiology fails on its own terms and is therefore explanatorily inadequate. Furthermore, as a reductive approach, the etiological approach is willing to abandon some of our pre-theoretic intuitions about teleological explanation. But my project is to give a naturalistic account of teleological explanation that saves the phenomena and hence vindicates our pre-theoretic intuitions.
1.4.2. Mechanism

1.4.2.1. Cybernetics

Historically the first mechanistic account of natural teleology is the cybernetics ‘regulatory’ approach. It starts with an account of purposiveness, that is, of what it is to be a purposeful or goal-directed system. The basic idea is that a system $S$ has a goal $G$ only if $S$ can attain and maintain a stable end-state $G$ by means of negative feedback from $G$ (Rosenblueth, Wiener and Bigelow 1943). A torpedo, for example, has the goal of hitting a ship only if the torpedo can maintain its trajectory towards the ship and hit it by means of negative feedback from the ship.

Negative feedback mechanisms involve a controlled system, a sensing-controlling subsystem and a property maintained within a certain range. When disturbing influences enter the system from the environment, the sensing subsystem detects deviations in the controlled property and modifies the system so that the property is maintained within the required range by means of a negative feedback causal loop from the controller. The goal $G$ is that state or condition that controls the trajectory of the system through negative feedback.

On the basis of this negative feedback account of what it is to be a purposeful system the cybernetic account derives an account of teleological explanation. The idea is that to say that $S$ did $x$ in order to $G$ just is to say that $S$ is a purposeful system and $S$ did $x$ because $G$ caused, through negative feedback, $S$ to $x$. For example, to say that the torpedo turned to the right in order to destroy the ship just is to say that the torpedo turned to the right because the ship moved to the right sending signals to the torpedo to change its direction accordingly. Let us now test this account against the desiderata.

Goals are part of the negative feedback mechanisms that explain the occurrence of the means. For goals to send signals back into the system they must be actual, occurring states. The cybernetic account thus avoids commitment to causation by non-actualia and hence satisfies (i). And since intentionality plays no role in the specification of the goals of the system or the role the play in explanation this account satisfies (ii). The cybernetic account does not posit intrinsically evaluative states of affairs. But it fails to capture the distinctive normative features of teleological explanations, thus failing to satisfy (iii). The cybernetic analysis specifies what goals are and what role they play in terms of mechanical concepts. Mechanical concepts
typically lack normative implications: if \(c\) mechanically causes \(e\) and there is a feedback loop between \(e\) and \(c\) such that \(e\) in turn partly causes \(c\), it does not follow that \(c\) ought to produce \(e\) or that \(e\) ought to produce \(c\), at least not in the normative sense of ‘ought’. Of course, insofar as causes necessitate their effects, \(e\) must occur if \(c\) does. But this ‘must’ refers to natural necessity not hypothetical necessity. For the same reason, the cybernetic account fails to meet the modal profile of hypothetical necessity. Finally, by virtue of taking teleological explanation to be a mechanical explanation by negative feedback, the cybernetic analysis fails to vindicate teleology as distinctive mode of explanation given the completeness of mechanism. So it fails to satisfy (iv).

That cybernetics fails to accommodate the normative and modal profile of hypothetical necessity is not surprising given that it mischaracterizes what goals are and what role they play (Jonas 1966). Except for fulfilled goals, goals are future unactualized state of affairs. Such states can’t send signals back into the system. For example, suppose that an animal is hungry and hence forms the goal of getting food and hence behaves appropriately by searching for food. Until the animal finds food there is no actual food to causally guide its behaviour. Yet, the animal has the goal. Or suppose there is no food to be had. It is still the case that the animal has the goal. The cybernetics analysis thus equivocates between ‘goals’ as the targets to which a goal-directed system is directed, and ‘goals’ in the proper sense of the state of affairs that would obtain if the system actually attains its target. For example, the ship is the target of the missile while hitting (destroying) the ship is the goal. The ship may be capable of sending signals back into the system and the system may be able to detect them. But the destroyed ship can’t.

To reconcile the unactualized character of a goal that has not yet been attained with its causal efficacy as a controller, a modification has been proposed to the cybernetic account: ‘\(S\) has goal \(G\)’ is true only if \(S\) can attain and maintain \(G\) by means of negative feedback from \(S\)’s representation of \(G\) (Adams 1979). Feedback loops could monitor whether the represented goal state, rather than the goal itself, is close to attainment and direct or modify the behaviour of the system accordingly. Of course, the notion of ‘representation’ calls for naturalization. But the problem with this modification is that ‘representation’ implies intentionality so this amendment reduces natural teleology to intentionality, thus contravening (ii).
By identifying goals with states that are part of negative feedback mechanisms, the cybernetic account locates goals in the causal structure of the world as per naturalism. But it fails to satisfy conditions (iii) and (iv) for teleology naturalized. And since goals figure as causes rather than as goals in the explanation of the occurrence of the means, the cybernetic analysis fails to accommodate the normative and modal profile of hypothetical necessity and hence it is explanatorily inadequate. So cybernetics offers an eliminativist account in which teleology just is an instance of a mechanical explanation by negative feedback.

1.4.2.2. Systems-Theory

According to the systems-theoretic account, ‘$S$ has a goal $G$’ is true only if $S$ behaves in a plastic and persistent way towards $G$, that is, only if $S$ tends to attain and maintain $G$ across a range of conditions (Bertalanffy 1969; Braithwaite 1964; Nagel 1961; 1977; Sommerhoff 1950).

‘Persistence’ or robustness is the capacity of the system to reliably attain $G$ starting from different initial conditions and from a number of alternative pathways by responding to its circumstances appropriately. ‘Plasticity’ or adaptiveness, in turn, is the capacity of the system to reliably maintain, or resume, its trajectory towards $G$ by making compensatory changes to perturbations. Biological organisms constitute the paradigmatic example of a goal-directed system. They possess characteristic homeostatic capacities that enable them to restore and maintain a variety of vital equilibrium states in the face of perturbations by making compensatory changes to their physiology, morphology and behavior (Ayala 1970). Specific homeostatic sub-systems such as the immune, thermoregulatory or endocrine systems that have the capacity to maintain a stable end-state by responding adaptively to perturbations are also paradigmatic examples of purposive systems. Purposiveness is thus an observable gross behavioural property of a system.

This account locates goals in the causal structure of the world and hence is naturalistically acceptable. A goal is just the end state to which a purposive system tends. It is thus an effect, but unlike the selected effects that constitute evolutionary functions, these effects don’t cause anything. Since goals are not causes the problem of causation by non-actualia does not arise, so this view satisfies (i). And since purposiveness is a gross behavioral capacity that is exhaustively specified in terms of plasticity and persistence rather than in terms of the intentions of an agent, this view also satisfies (ii). Also, this account does not posit intrinsically evaluative states of
affairs to individuate goals. Rather goals are those states of affairs that a system can approach in a plastic and persistence way. However, there is no sense in which goals require the means such that we can evaluate whether the means is good for attaining them. In fact, this account doesn’t assign a specific explanatory role to goals. Rather, it is the overall disposition of the system to attain (and maintain) a set of equilibrium states that explain the behavior of the system. So it fails to satisfy condition (iii). Arguably, dispositional explanations are causal explanations. So it also fails to satisfy condition (iv).

Now, plastic and persistent behavior exhibits the characteristic modal profile of hypothetical necessity. A purposive system is a system that reliably attains a set of equilibrium states across a wide range of conditions. That is, it is a system that can attain the same effect across different causal pathways such that, if this particular causal pathway were perturbed, the system would resume its trajectory through an alternative pathway. This means that the system would do (ceteris paribus) what is (hypothetically) necessary to attain the relevant set of equilibrium states. However, the systems-theoretic account is reductionist. It seeks to specify what it is to be a purposive system in non-circular strictly behavioral terms. So it lacks the conceptual resources to capture the modal and normative character of hypothetical necessity as such. On account of this it is explanatorily inadequate.

1.4.2.3. Autonomous Systems

The latest edition of mechanism offers a naturalistic account of purposiveness in terms of the distinctive causal organization of ‘autonomous’ systems. The basic idea is that a system \( S \) has a purpose \( G \) if and only if the system must produce \( G \) to maintain itself (Barandiaran et al. 2009; Christensen & Hooker 2002; Di Paolo 2005; Mossio et al., 2009; Moreno and Mossio 2015; Thompson 2007). Purposive systems are ‘autonomous systems’. These are non-linear, far from equilibrium systems that maintain its internal organization by adaptively regulating its dynamical coupling with the environment. A paradigmatic example is a metabolic system. A metabolic system is constantly exchanging energy and matter with the external environment allowing the system to dynamically maintain its organization by means of a network of processes that depend recursively on each other. Like its predecessors, the autonomous-systems analysis derives an account of teleological explanation from this account of purposiveness. The idea is that to say that \( S \) did \( x \) in order to \( G \) just is to say that \( S \) is a purposive system and \( S \) did \( x \) because
$G$ is necessary for $S$’s self-maintenance. For example, to say that a cell produced a certain amount of metabolites in order to metabolize just is to say that the cell produced a certain amount of metabolites because doing so is necessary to maintain itself.

This view of teleological explanation is clearly naturalistic. Goals are the stability-states that the system must attain in order to exist. So goals are located in the causal structure of the world, namely, in the distinctive organization of living beings. And like the systems-theoretic account but unlike the etiological and cybernetic account, this view does not conceive goals as causes. So this account avoids commitment to causation by non-actualia and thus satisfies condition (i). Furthermore, no assumption is made to the effect that goals are constituted as goals by virtue of being represented by the system as to be attained. So this account avoids reducing natural teleology to intentionality and hence satisfies condition (ii). However, explanations in terms of self-maintenance are mechanistic explanations. After all, an autonomous organization is a network of causal-mechanical processes that depend recursively on each other. So this account does not satisfy (iv).

A distinctive feature of this account is that unlike the cybernetics and the system-theoretical account it explicitly aims to naturalize the normativity of teleological explanation. According to this view, because the stability states with which goals are identified must occur for the system to exist, the system ought to produce them. These conditions of viability thus specify the norms that the system must conform to in order to exist. Activities that fail to contribute to the maintenance of the overall stability of the system and hence to conform with these norms are thus not doing what they are supposed to do. In this sense this account claims to capture the distinctive normative features of teleological explanation without positing intrinsically evaluable state of affairs and without appealing to the historical process of evolution by natural selection.

But there are two problems. First, even accepting for the sake of the argument that being necessary for self-maintenance is a genuinely normative relation, it is not the kind of normative relation that holds between a goal and its means. The former is a relation of causal necessity while the latter is a relation of hypothetical necessity. So the former lacks the modal profile

---

1 Like the etiological approach the autonomous-systems approach takes functional explanations to be clear even paradigmatic instances of teleological explanation (Mossio et al. 2009; Moreno and Mossio 2015).
characteristic of teleological explanation. In teleological explanation means are hypothetically rather than causally necessary for attaining the goal. The idea is that if this particular means had failed to obtain some other would have obtained to attain the goal across a range of conditions. So the particular means is causally sufficient but not necessary for the goal. But on the autonomous-systems approach means are causally necessary for attaining the goal. So nothing follows about what the system would have done in counterfactual circumstances. This account thus holds for states or effects that are goals as well as for those that are not. So this view takes too many things to be goals.

The second problem is that there are many things that a purposive system ought to do, for example, to meet some biological goal, that are not pre-conditions for the system’s continued persistence. Purposive systems typically can attain states that constitute malfunction without thereby ceasing to exist. Rather, the system continues to exist but does so at a less than optimal level. Furthermore, the activities $x$ of $S$ in these cases do not count as means to ends because the ends have not been attained. But on the present account, these activities could still be explained in the same way that genuine means are. Normativity is thus an all-or-nothing thing: failure implies non-existence. However, an animal can have the goal of eating an insect and the thermoregulatory system may have the goal of slightly lowering an animal’s temperature. Neither will die if they fail to achieve their goals. So failure to do what the system ought to in order to attain some goal does not imply ceasing to exist. So this view also takes too few things to be goals. So the account fails to accommodate the normativity of teleology and hence fails to satisfy (iii).

In conclusion, the autonomous system account offers an important account of the kind of organization that realizes purposiveness in the causal structure of the world. But by identifying goals with the stability states that are necessary for the self-maintenance of an autonomous system it fails to accommodate the distinctive normative and modal profile of teleological explanation.

1.4.3. Normativism

A completely different strategy from that of etiology and mechanism is the normative approach. This approach starts with an account of the concept of a goal rather than the concept of
purposiveness and argues that given what goals are they figure in a distinctive form of explanation, teleological explanation. On account of this, I will argue, this approach is explanatorily adequate. The two main normative accounts share the view that the concept of a goal is an irreducibly normative concept. But they differ about what the relevant normativity is supposed to be. Because of its irreducibly normative character, this approach takes the concept of a goal to be primitive, hence not analyzable in non-normative scientific terms. On account of this the normative approach fails to satisfy naturalism.

1.4.3.1. **Axiology**

Like the etiological approach the axiological or value approach is explicitly meant to capture the distinctive features of teleological explanations. But unlike the etiological account it does so by capturing what constitutes a state of affairs as a goal. The idea is that ‘S did x in order to G’ is true if and only if S did x because [S’s x-ing contributes to the attainment of G and G is good] (Bedau 1991; 1992). A state of affairs G constitutes a goal rather than a mere stable-end-state if and only if G is good. G’s being good, in turn, imposes a demand or a normative requirement on S to bring about the means x to attain G and hence explains why S did x. Teleological explanations thus have an irreducible evaluative content. For example, to say that she ran in the morning in order to exercise is to say that she ran in the morning because running is a means to exercise and exercise is good. And to say that the immune system produced a certain amount of lymphocytes in order to restore the health of the organism is to say that the immune system produced that amount of lymphocytes because it is a means to restoring the health of the organisms and restoring health is good. To explain the occurrence of some event teleologically is thus to explain it as being good. The goodness of the goal on account of which it explains is an intrinsic property of that state of affairs.

Bedau (1992) distinguish three grades of evaluative involvement in a teleological explanation:

1. \( S \) does \( x \) in order to \( G \) iff \( S \) does \( x \) and \( S’s \ x-ing \) contributes to the attainment of \( G \) and \( G \) is good for \( S \).

2. \( S \) does \( x \) in order to \( G \) iff \( [S \ does \ x \ because \ S’s \ x-ing \ contributes \ to \ attaining \ G] \) and \( G \) is good.
3. $S$ does $x$ in order to $G$ iff $S$ does $x$ because [$S$’s $x$-ing contributes to attaining $G$ and $G$ is good].

He argues that only grade 3 constitutes a genuine teleological explanation on the grounds that only by including the value condition in the scope of the ‘because’, rather than just its consequences, allows us to capture the fact that goals explain by virtue of being good. The problem with the etiological, cybernetic and the systems-theoretic accounts, Bedau argues, is thus that they take the ‘because’ to have narrow scope and hence include only the consequences but not the beneficiary in the content of the explanation. On account of this they assign a mere causal role to goals. But the explanatory role of goals is irreducibly normative. So the axiological account shows how teleology can explain something that mechanism can’t. So this account satisfies condition (iv): Teleological explanations are a distinctive mode of explanation by virtue of its irreducibly normative content.

Since goals do not explain the occurrence of their means by causing them this account avoids commitment to causation by non-actualia and hence satisfies (i). And since the goodness of the goal is an intrinsic property of that state of affairs (more on this below), the goodness does not depend on the system representing that state of affair as good. So this account avoids reducing natural teleology to intentionality and hence satisfies (ii). Also, by virtue of being good the goal $G$ requires the system $S$ to bring about the appropriate means to its attainment. So this account captures the normative profile of hypothetical necessity. However, it does so by positing intrinsically evaluative states of affairs. So it fails to satisfy condition (iii). Goodness, on this view, is a primitive feature of the world that cannot be accounted for in other non-normative terms. Bedau offers no account of the conditions that realize goodness in the causal order of the world. On account of this the axiological account fails to be naturalistically acceptable.

Against the view that the normativity of teleology requires the goodness of goals, Walsh (2008) distinguishes between two normative relations, ‘ought’ and ‘normative requirements’ (Broome 1999). The idea is that a system is normatively required to bring about the means to attain its goals whether or not the goals ought to be attained. For example, an agent might have the goal of committing suicide. There is, I take it, nothing intrinsically good about committing suicide. Nevertheless, it is still the case that given that goal the agent is normatively required to bring about the appropriate means to fulfill that goal. So if drinking poison is the appropriate means to
attain that goal then drinking poison is good for committing suicide. This is true even if, arguably, one ought not to commit suicide. As this example shows, because the goal of suicide normatively requires drinking poison, the goal of dying explains why the agent drank poison, even if the goal is not in itself good. Or consider a more trivial example. Suppose there is no fact of the matter whether one ought to have coffee with milk or black. But given that one wants coffee with milk one is normatively required to pour the liquid into the mug. So again, the normativity of hypothetical necessity does not require the categorical goodness of goals. This distinction allows us to vindicate the normative relation between goals and means without being committed to the goodness of the goal.

This account focuses on the normative profile of hypothetical necessity but is silent about its modal profile. On the assumption that these are two aspects of the same thing, satisfying the normative aspect implies that the modal aspect is also satisfied, at least implicitly. So the axiological account is explanatorily adequate. Teleological explanations constitute a distinctive mode of explanation.

1.4.3.2. Rationality

According to the rationality account, to say that ‘S did x in order to G’ is to say that S’s doing x was directed at G, and S’s x-ing towards G is part of a rational interpretation of the behavior of S (Sehon 2005). The idea is that teleological explanations play an indispensable role in the interpretation of the behavior of intentional agents. Interpretation, in turn, is constituted by the principle of rationality. This principle dictates that to explain an action as the consequence of the intention of an agent is to show that the agent was rationally required to do x given G (Davidson 1963). To act for a goal or reason is thus constitutive of intentional agency and teleological explanations capture this fact (more on this in Chapter 6).

Since goals (reasons) explain by making rational sense of the behavior of an agent rather than by causing the behavior, the rationality account shows that teleological explanation is distinctive mode of explanation from causal-mechanical explanations, so it satisfies (iv). And for this reason, it also avoids commitment to causation by non-actualia, so it satisfies condition (i). However, this account applies only to intentional agents, that is, agents subject to the norms of practical and theoretical reason that have cognitive and conative mental states. So this account
reduces teleology to intentionality and hence fails to satisfy condition (ii). The problem is that although some non-human organisms can be interpreted as rational agents not all can. So if organisms can be interpreted as purposive systems even if not as rational agents this account excludes them. Furthermore, although the endocrine or immune system—or any other homeostatic systems for that matter—are not intentional agents they can be explained teleologically.

If $S$ has goal $G$ and doing $x$ is the means to attain $G$ under the relevant conditions $C$ then $S$ ought to or has reason to or is required to bring about $x$. The agent is thus justified in acting in that way. According to Sehon, rational interpretation provides two conditions for normative evaluation: First, it provides a measure of the degree to which the explanation renders the behavior to be explained as *appropriate* for achieving the goal, and second, it provides a measure of the degree to which the goal is a *valuable* state of affairs. And since a state of affairs constitutes a goal just in case there is reason to pursue it, goals explain the occurrence of actions as goals. So this account captures the normative profile of hypothetical necessity. And unlike the axiological account, the fact that $S$ is normatively required to bring about $x$ is not grounded in the fact that $G$ is intrinsically good. Rather, it is grounded in the fact that teleological explanations are a constitutive part of rational interpretation. So the rationality account captures the distinctive normative features of teleological explanation without positing intrinsically evaluative states of affairs. In this respect this account satisfies condition (iii).

Furthermore, an essential feature of rational agents is their rich behavioral flexibility that allows them to respond in an open-ended way to their circumstances. The modal profile of this behavioral flexibility is precisely that of hypothetical necessity: A rational agent would do what is necessary for attaining her goal across a wide range of conditions. So the rationality account captures the distinctive modal profile teleological explanations. On account of this this view is explanatorily adequate.

However, there is no indication of how rationality is causally realized in the world. Instead, rationality is a primitive principle in a theory of interpretation. So this account fails to satisfy naturalism despite avoiding positing intrinsically evaluate state of affairs.
1.5. Conclusion: A Dichotomy between Eliminativism and Primitivism

The general strategy of mechanism (and etiology) is to start with a naturalistic account of what it is to be a purposive system and then to ground an account of teleological explanation on that basis. This approach gives variously good accounts of what it is to be a purposive system. But it fails as an account of teleological explanation. On account of these they are naturalistically acceptable at the expense of being explanatorily inadequate. In turn, both normative accounts of teleological explanations vindicate the distinctive explanatory role that goals play as goals in explanation and hence are explanatory adequate. But they take the normativity of goals to be a primitive feature of the world and hence they are silent about the place of goals in the causal structure of the world. For this reason they fail to be naturalistic.

The dialectic between etiology and mechanism on the one hand, and normativism on the other exhibits the following pattern: While the former eliminates the distinctive role that goals play in explanations in order to vindicate their naturalistic status, the latter vindicates the distinctive explanatory role of goals but at the expense of failing to show their place in nature. This, I submit, generates a dichotomy between eliminativism and primitivism respectively.

Eliminativism is the view that goals are part of the natural order but they play no explanatory role as goals. Hence, eliminativism is the view that teleological explanation is naturalistic by virtue of being really an instance of some form of causal-mechanical explanation. Primitivism on the other hand is the view that goals play a distinctive explanatory role as goals but is silent about their place in the causal structure of the world. So primitivism is the view that teleological explanation is a distinctive form of explanation by virtue of its irreducibly normative character.

Teleology, it seems, can’t be both, natural and distinctively explanatory at the same time. This result is not a problem for the standard approaches given their explicit aims. Of course, a defender of mechanism can say ‘so much for the distinctive explanatory role of goals’, the normativist can say ‘so much for locating goals in nature’.

For this reason, none of the standard approaches meets my explicit aim to satisfy both desiderata at the same time. This is not to say that in looking for an alternative account that avoids this dichotomy I will not make use of these approaches. The next task, however, is not yet to give an
alternative account but to offer a diagnosis that will expose the underlying assumptions shared by these views that generate the dichotomy.
Chapter 2  
Two Dogmas of Naturalism

2.1. Introduction

In the previous chapter I introduced the project of giving an account of teleological explanation that is both naturalistically acceptable by virtue of showing the place of goals in the causal structure of the world, and explanatorily adequate by showing that goals play a distinctive explanatory role as goals. Then I tested the standard etiological, mechanical and normative approaches against these desiderata. I argued that while the etiological and the mechanical approaches satisfy naturalism at the expense of explanatory adequacy, the normative approach satisfies explanatory adequacy at the cost of failing to satisfy naturalism. This dialectic, I concluded, generates a dichotomy between eliminativism and primitivism. Eliminativism is the view that goals are part of the natural causal structure of the world but not distinctively explanatory as goals. Teleological explanations are thus nothing but an instance of causal explanation. Primitivism, in turn, is the view that goals play a distinctive normative role as goals in explanation but it offers no account of goals as part of the natural order. Teleological explanations are thus irreducibly normative. The project I envisage, however, requires a middle way between these two extremes. So an account of why this dichotomy is generated and hence of why the standard approaches fail to satisfy both desiderata at the same time is called for.

In this chapter I offer a diagnosis of this dichotomy. I argue that the dichotomy is predicated on a set of stringent assumptions about what the method of naturalism requires shared by the standard approaches. Naturalism requires that goals be located in the causal structure of the world by indicating the conditions that causally realize them. But the standard approaches add two assumptions. One assumption concerns the individuation of goals. The idea is that to say what a goal is just is to say how it is causally realized. The other is about how goals figure in explanations. The idea is that the role that goals play in explanation can be specified exhaustively by citing the way they are realized. These assumptions are not restricted to the standard approaches to teleological explanation. Rather, they are pervasive in naturalistic philosophy. For this reason I call them two dogmas of naturalism. Roughly, these are:
D1. To say what some phenomenon X is just is to say how X is caused or realized.

D2. The theoretical role of some phenomenon X can be specified exclusively by citing the way X is causally realized.

D1 concerns the individuation of theoretical entities. For example, to say what water is or whether some stuff is a member of that kind just is to say how water is causally realized, namely, by the chemical composition H2O. In a slogan, to be is to be causally realized or individuation is causal realization. D2, in turn, is about how these entities figure in explanations. For example, the theoretical role that water plays in chemistry can be completely specified by citing its chemical structure such that everything that the concept of water explains can be explained without loss of information by citing its composition. In a slogan, explanation is causal realization. D2 follows from D1 and the independently plausible principle that to say what something is just is to say what its theoretical role is, or alternatively, that the theoretical role of something is exhausted by its individuation conditions. My aim in this chapter is not to refute these dogmas but to identify and articulate them and to show how they impose a choice between eliminativism and primitivism. Also, I suggest that this dichotomy generalizes beyond the problem of teleological explanation whenever normative concepts are involved, such as ‘mental content’, ‘knowledge’, and ‘action’.

The structure of the chapter is as follows. In section 2.2 I introduce D1 and go through some examples of contemporary philosophical theories that are committed to D1. In section 2.3 I introduce D2 and offer some examples. Then, in section 2.4 I argue that eliminativism is generated by the commitment of mechanism to these dogmas and in section 2.5 I argue that primitivism is generated by the rejection of the normative approach of the two dogmas. Finally, in section 2.6 I conclude that although the standard strategy of naturalization is committed to the two dogmas, these are further theoretical commitments not entailed by naturalism itself. So although the naturalist is free to appeal to this dogmatic strategy of naturalization, naturalism per se does not necessitate it. The naturalist is thus free to employ an alternative non-dogmatic hence less stringent strategy of naturalization in order to satisfy both, naturalism and explanatory adequacy.
2.2. The First Dogma of Naturalism: Individuation is causal realization

According to the first dogma (D1), to say what some phenomenon X is just is to say how X is caused or realized. The idea is that an account of how some phenomenon is causally realized provides a constitutive analysis of what that phenomenon is. So the properties, conditions or mechanisms that cause or realize a phenomenon are not only causally sufficient for their natural occurrence, they also determine its identity and hence provide a principle of individuation for that phenomenon. For example, to say what water is just is to say how water is causally realized, namely, by the chemical composition H2O.

As we saw in the first chapter naturalism is the view that the causal-physical-order of the world as described by natural science exhausts reality. It follows that if we are committed to the existence of some entity and hence we are realist about that entity we need to locate that entity in the causal structure of the world as described by natural science (Jackson 1998). But what does locating an entity in the causal structure of the world consists of? I take it that to show that X is part of the causal structure of the world requires showing how is X causally realized. This is an empirical issue that implies nothing by itself about what X is or about what role X plays in explaining some other phenomenon. But according to the standard method of naturalization, to locate some putative sui generis phenomenon is precisely to give a constitutive analysis of what that phenomenon is in non-circular strictly causal-mechanical terms. So the standard method of naturalization is committed to D1: A theoretical entity X is individuated by its causal realizer. Fodor (1990, 32) gives a classical formulation of this strategy of naturalization in the context the problem of naturalizing intentionality:

“The worry about representation is above all that the semantic (and/or the intentional) will prove permanently recalcitrant to integration in the natural order [...]. What is required [...] at a minimum, is the framing of naturalistic conditions for representation. That is, what we want at a

---

2 Jackson takes location to consist in what he calls “entry by entailment thesis”. The idea is that “the one and only way of having a place in an account told in some set of preferred terms is by being entailed by that account” (p. 5). By “entailment” he means “simply the necessary truth-preserving notion—call it ‘necessary determination’ or ‘fixing’ if you prefer.” This ‘fixing’ is then cashed out in terms of a supervenience relation.
minimum is something of the form “$R$ represents $S$” is true iff $C$ where the vocabulary in which condition $C$ is couched contains neither intentional nor semantic expressions.”³

So according to Fodor naturalizing intentionality requires not just showing how representation is causally realized but also identifying representation with its causal realizer. Locating intentionality in the natural order thus takes the form of giving a constitutive analysis of intentionality in terms of its place in the natural order. A constitutive analysis of some phenomenon specifies a relation of metaphysical determination between the phenomenon to be naturalized and the set of necessary and sufficient conditions that the phenomenon has to satisfy to be the kind of thing that it is. The extensional adequacy of the analysis is then tested against intuitions for potential counterexamples, which are typically identified by means of thought experiments. But since this method of naturalization is explicitly reductionist, there is no demand to preserve all our pre-theoretic intuitions about the phenomenon in question. For example, intentional states are pre-theoretically taken to underwrite rational norms that impose certain demands on what an agent ought to do. For example, given that Lupe desires to buy some groceries and that she believes that they can be found at the local store, Lupe ought to go to the local store. But defenders of the standard approach to naturalizing intentionality downplay the importance of this normativity. For example:

“The trouble is that requiring that normativity be grounded suggests there is more to demand of a naturalized semantics that it provide a reduction of such notions as, say, extension. But what could this “more” amount to? To apply a term to a thing in its extension is to apply the term correctly; once you’ve said what it is that makes the tables the extension of “table”’s, there is surely no further question about why its’ correct to apply a “table” to a table. It thus seems that if you have a reductive theory of the semantic relations there is no job of grounding normativity left to do.” (Fodor 1990, 135, footnote 35)

And Dretske (1994, 482 footnote 16) concurs:

³ Later on Fodor (1990, 96) has modified this by demanding just sufficient conditions specifiable in naturalistic terms rather than necessary and sufficient conditions.
“I agree with Fodor that the only normative quality a naturalistic theory of meaning has to explain is the quality of being able to mean something that isn’t so. If we can solve the problem of misrepresentation—or equivalently, the disjunction problem—we will have all the normativity we want.”

After all, the natural relations with which representation is identified are generally taken to be non-normative. The problem for this strategy is of course that if we take this normativity to be an essential feature of intentionality the reduction collapses into elimination. The lesson is that by virtue of its commitment to D1, the standard strategy imposes quite stringent demands on naturalization: To naturalize a phenomenon X is to identify X with the conditions that causally realize X. Again, Fodor (1987, 97) gives a classical statement of the strong reductionist implications of the standard strategy applied to the case of naturalized intentionality:

“I suppose that sooner or later the physicists will complete the catalogue they’ve been compiling of the ultimate and irreducible properties of things. When they do, the likes of spin, charm, and charge will perhaps appear on the list. But aboutness surely won’t; intentionality simply doesn’t go that deep. It is hard to see, in face of this consideration, how one can be a Realist about intentionality without also being, to some extent, a Reductionist. If the semantic and the intentional are real properties of things, it must be in virtue of their identity with (or maybe of their supervenience on?) properties that are themselves neither intentional nor semantic. If aboutness is real, it must be really something else.”

The pervasiveness of this dogma in contemporary naturalistic philosophy is manifested in the plethora of causal theories of X that aim to be constitutive accounts of X and hence to provide a causal principle of individuation. I will consider the case of the causal theory perception, content and knowledge. My aim is not criticize these views but only to make explicit their commitment to D1. Also, I won’t argue here that these causal theories face a dichotomy between eliminativism and primitivism. But I believe this dichotomy generalizes to any account of normative concepts that aim to be specified in non-normative terms. Let us consider some examples of D1 in philosophy.

---

4 The primitivist response to the eliminativist excesses of this standard approach to intentionality naturalized can be found in McDowell (1994) and Brandom (1994).
2.2.1. The Causal Theory of Perception

In having a perceptual experience something is conveyed to the subject. Perceptual experiences are typically assumed to have representational content and to be (at least partly) individuated by their representational content. The representational content of a perceptual experience can be thought of as given by the conditions under which it is accurate. What a perceptual experience conveys to the subject is that those conditions are satisfied (Siegel 2015). According to the causal theory of perception (CTP), when $S$ perceives $O$, $S$ forms an experience with the complex (existentially quantified) content of the form: There is an $O$ at location $L$, and the fact that there is an $O$ at $L$ is causing this experience. Notice that two things are included in the content of a perceptual experience according to CTP: the experience itself (self-referential content) and the causal relation between the experience and what causes it. Thus, CTP is not just the view that for $S$ to perceive $O$ is for $O$ to cause (in the right way) $S$’s perceptual experience. It is also the view that $O$’s causing $S$’s perceptual experience figures in the content of $S$’s perceptual experience.

The claim that perception of an object $O$ requires causal contact with $O$ is a perfectly plausible empirical condition for perception of $O$ to be possible. But since this is a causal and hence a metaphysically contingent claim about perception, it entails nothing about what the content of the perceptual experience is and hence about how perceptual experiences are individuated. In particular, it does not entail that perceptual experiences are individuated by the way they are caused or that their content is or includes the fact that the relevant causal relation holds. But of course CTP is not expressing an empirical contingent truth about a necessary causal condition for perceptual experience to occur. Rather, CTP is making a constitutive claim about the nature of perceptual experience according to which, having a perceptual experience of an object consists of the fact that the object is causing the experience. And this stronger claim does entail that the perceptual experience is individuated by what caused it and that the conditions that cause the perceptual experience figure in the content of the perceptual experience. So according to CTP, to say what a perceptual experience is just is to say how the perceptual experience is caused. Hence CTP is committed to the first dogma D1.
2.2.2. The Causal Theory of Mental Content

Consider now the causal theory of mental content (CTC). According to CTC a mental representation $M$ represents that $P$ if and only if $P$ causes $M$ (in the right way)\(^5\) (Fodor 1984; 1992). The idea is that under normal conditions our causal encounters with the world are what make it the case that our thoughts have the contents that they do. For example, the mental representation of the concept DOG represents dogs because causal interactions with dogs cause us under normal conditions to form that representation. So the way a mental representation is normally caused determines what the representation is about. Since mental representations are individuated by their content, it follows that mental representations are individuated by the way they are normally caused as per D1. Unlike CTP however, CTC does not entail that the content of the thought is that there is a dog and that the fact that there is a dog is causing the thought. Our dog-thoughts are about dogs not about the fact that dogs cause us to think about them.

Notice again that although having causal encounters with dogs is an empirically plausible causal condition for thinking about them, it doesn’t follow that thinking about dogs consists in or is constituted by causal encounters with them. The point is not just the trivial one that one can think about dogs in the absence of an ongoing encounter with dogs. Rather the point is the substantial one that to say what having a thought about something is, just is to say how thoughts are normally caused such that thoughts are individuated by the way they are normally caused. The former is an empirical contingent truth about a necessary causal condition for dog-thoughts to occur. But the latter is a conceptual/metaphysical claim about the nature of mental content, that is, about a metaphysically constitutive condition for being a dog-thought. The former implies the latter only if the former is taken to be an analysis of the latter. So by virtue of being an analysis CTC entails that the content of a mental state is what causes that state.

---

\(^5\) This proviso is endemic to many causal accounts. I take it to indicate the poverty of causal accounts to reconstruct or translate concepts that are characteristically normative in non-normative terms. There are of course a variety of proposals about how to cash what ‘the right way’ amounts to. For example, some prominent options are in terms of normal conditions (Stampe 1977), evolutionary conditions (Millikan 1991), developmental conditions (Dretske 1988), and asymmetric dependence relations (Fodor 1990).
As mentioned before, it is not surprising that defenders of CTC down play the normative character of content. It is not clear that the causal relation with which content is identified can capture the relevant normative implications. In that case, CTC faces the problem of elimination.

2.2.3. The Causal Theory of Knowledge

Finally, let’s now look at the causal theory of knowledge (CTK). CTK claims that a subject $S$ knows that $P$ if and only if $P$ is true; $S$ believes that $P$; and $S$’s belief that $P$ is caused (in the right way) by the fact that $P$ (Goldman 1967; 1976). The basic idea is that in order to have knowledge that $P$ a subject must believe that $P$, $P$ must be true and the belief must be caused by the very fact that makes it truth, that is, $P$. In other words, knowledge requires a causal chain between a subject’s belief in some proposition and the truthmaker of that proposition.

Again, it seems like an empirically correct and philosophically uncontroversial thing to say that, at least in the case of empirical knowledge or knowledge about contingent matters of fact, knowledge requires some causal connection between the subject and the fact known. After all, empirical matters of fact are essentially causal facts and hence facts that belong in the causal order of the word. One cannot therefore know that $P$, where $P$ is an empirical matter of fact, unless one is causally located with respect to $P$, that is, unless one is causally connected with $P$ in some sense (e.g. perceptually or by testimony). But from the fact that for empirical knowledge to be causally possible there must be some broadly conceived causal connection between the knower and the fact known, it doesn’t follow that it is a constitutive feature of what knowledge of empirical matter is that the connection must hold. But by virtue of constructing this causal account as a constitutive analysis of knowledge, CTK makes the causal connection a conceptual or metaphysical necessary feature of what knowledge is. Hence, CTK is committed to D1.

The concept of knowledge is usually taken to be a normative concept. After all, the concept requires justification and justification is normative. So by virtue of its exclusive causal character,

---

6 Despite its original motivation, CTK has been subject to Gettier-style counterexamples in which the relevant causal connection between the belief and the fact that makes it true holds but where the subject forms the belief by luck. Although I don’t intend to go into these counterexamples, I do want to address a crucial aspect of CTK exposed by these counterexamples, namely, the alleged constitutive relation between knowledge and causation.
it is not clear that CTK can capture the normative content of attributions of knowledge. In that case it seems that CTK leads us into eliminativism about knowledge.

2.2.4. Summary and Evaluation

All these causal theories share a common motivation and a common methodological strategy. The motivation is naturalism. The idea is that the normative character of perception, intentionality or knowledge imposes a demand of showing how these phenomena fit within the naturalistic picture of the world. The methodological strategy is that of accounting for what perception, content and knowledge are in terms of what causes them. That is, all these theories conform to the pattern according to which first it is identified how the phenomenon in question is caused and then, such an account is made into a constitutive analysis of what that phenomenon is, its principle of individuation as per D1.

I submit that by construing naturalization as a constitutive analysis, this strategy runs together two logically independent questions. The first is what we might call the causal question, namely, how is some phenomenon possible in the causal order of the world? This question is a demand for showing the mechanisms or properties that cause or realize that phenomenon. And the second question is what might be called the constitutive question, that is, what makes such a phenomenon the phenomenon that it is? This question is a demand for a principle of individuation that determines the essence and hence the identity of the relevant phenomenon. What makes these questions logically independent is that the answer to the first one implies nothing, by itself, about the answer to the second. There may be phenomena for which this inference holds. But no reason has been giving to think that all phenomena must conform to this inference. Hence the dogmatic character of the assumption.

2.3. The Second Dogma of Naturalism: Explanation is Causal Realization

According to D2 the theoretical role of a phenomenon X can be specified exclusively by citing the way X is causally realized. That is to say that for any theoretical category, X, X's explanatory role is exhaustively accounted for by the way that X is causally realized. D2 follows from D1 and the independently plausible principle that to say what something is just is to say what its
theoretical role is, or alternatively, that the theoretical role of something is exhausted by its individuation conditions. The idea behind D2 is that the theoretical role of a phenomenon is exhausted by the causal realizer of that role such that the former can be completely specified in terms of the latter. Or alternatively, D2 is the claim that to say what the content of an explanation is just is to say what the underlying causal conditions for its realization are. So for example, the chemical composition of water is H2O. As a theoretical entity the chemical composition of water is taken to provide the individuation conditions of water as per D1. Having identified water with its chemical realizer, it follows that everything that water explains can be explained in terms of the causal realizer of water. Thus, water figures in chemical generalization as H2O. So as per D2, the theoretical role that water plays in chemistry can be completely specified by citing its chemical structure. In a slogan, to explain is to causally realize or explanation is causal realization. Let us now consider some more examples.

The idea behind these examples is again not to criticize them but to expose the role of D2 in each case. However, I will point out how the eliminativist consequences of specifying the theoretical role of some phenomenon X exclusively by citing the causal realizer of X can be problematic where ‘X’ plays a normative role.

2.3.1. The Molecular Gene

Genes play a crucial theoretical role in biology. They are the basic units of inheritance and development. Genes are realized by linear sequences of DNA molecules each consisting of a double chain of nucleotides. But before the discovery of the molecular structure classical genetics conceived of genes as the functional units in chromosomes whose differences cause phenotypic differences (Waters 1994). The conditions that causally realize functional genes were unknown and hence the theoretical role of the functional concept of a gene was not specified in terms of its molecular realizer. Classical genetics was thus not committed to D1 or to D2.

In contrast, contemporary genetics conceives of genes as molecular units of DNA that function to determine linear sequences of nucleotides. Genes, on this view, serve as templates in the synthesis of RNA molecules, which in turn function as templates in the synthesis of polypeptide molecules. Synthesized polypeptides are linear sequences of amino acids that constitute proteins, which play many different functional roles in cells and organisms. This concept of a gene is
defined in molecular terms and hence in term of its causal realizer (Waters 2000). So molecular genes are individuated by their causal realizers as per D1. Having identified genes with their molecular realizers, the theoretical role that genes play in development and inheritance can then be exhaustively accounted for by specifying their molecular realizers as per D2.\footnote{There is of course considerable controversy as to which set of molecules at which stage of the transcription process constitutes genes, which partly underlies the controversy whether classical genetics can be reduced to molecular genetics (see Kitcher 1984; Schaffner 1993; Dupré 1993).}

The basic molecular concept, according to this analysis, is the concept of a gene for a linear sequence in a product of DNA expression: A gene $g$ for linear sequence $l$ in product $p$ synthesized in cellular context $c$ is a potentially replicating nucleotide sequence, $n$, usually contained in DNA, that determines the linear sequence $l$ in product $p$ at some stage of DNA expression (Waters 2000).

Notice that there is nothing wrong in itself about being committed to D1 and D2 in genetics. In particular, there is nothing wrong about saying that genes are nothing but the molecules that realize them or that everything that genes explain can be explained in terms of their molecular realizers. After all, the biological generalizations in which genes figure lack normative implications. Instead, if the molecular analysis is right it seems that having discovered how genes are causally realized not only allows us to precisely locate them in the causal structure of the world but also allows us to determine their identity and theoretical role.

2.3.2. The Causal Theory of Action

In contrast consider the phenomenon of intentional explanation.\footnote{I give an extended discussion the causal theory of action in Chapter 6.} We typically explain why rational, psychological agents act in the way they do by citing their goals, purposes, reasons, intentions or values. For example, we say that agent $S$ did $x$ in order to $G$, where $G$ is a goal or purpose that $S$ intends (or desires) to attain. Intentional explanations are, at least on the surface of it, teleological. But one can, of course, give an (in principle complete) explanation of how the behavior was causally produced by citing the way the neurophysiological mechanisms that causally realize mental states produce the relevant behavioral effects. One might also think that these two kinds of explanation, teleology and mechanism, form a nice methodological and conceptual division of labor: While mechanism explains the occurrence of the action as a
physical process (movement), teleology explains it as an action (intentional behavior) by citing the desires an intentions of the agent. As we saw in section 1.3.4 of the first chapter, the problem with this harmonious compatibility is however the completeness of mechanism, the view that given the causal-mechanical explanation, there is no further fact to be explained. The completeness of mechanism threatens to make teleological-intentional explanation redundant. This problem is nicely expressed by Malcolm (1968, 52-53). He invites us to picture a man that climbs up a ladder to retrieve his hat:

“The envisaged neurophysiological theory was supposed to provide sufficient causal explanations of behavior. Thus the movements of the man on the ladder would be completely accounted for in terms of electrical, chemical, and mechanical processes in his body. This would surely imply that his desire or intention to retrieve his hat had nothing to do with his movement up the ladder. It would imply that on this same occasion he would have moved up the ladder in exactly this way even if he had no intention to retrieve his hat, or even no intention to climb the ladder. Given the antecedent neurological states of his bodily system together with the general laws correlating these states with the contraction of muscles and the movement of limbs, he would have moved as he did regardless of his desire or intention.”

Not surprisingly, in order to avoid the elimination of intentional explanation in the face of the completeness of mechanism, the standard response has been to reduce intentional explanation, which has a distinctive teleological content, to causal explanation. According to the standard version of the causal theory of action (CTA), ‘S did x in order to G’ if and only if S did x because S wants G and believes that x is the means to G and these intentional states cause (in the appropriate way) x to occur (Davidson 1963). The idea is that to act is to respond to the way the agent represents the world as being (beliefs) and to the way the agent would want it to be (desires). Actions are generated and hence explained by the mechanical interaction of representational states. The general idea is that the reason is constructed by the brain state and in virtue of its causal constitution causes the behavior (Dretske 1988; Mele 2003). So having identified intentional states and hence reasons with the brain states that causally realize them as per D1, this theory takes the role that intentional states play in the explanation of action to be exhaustively specified in terms of their realizers as per D2.
2.3.3. **Summary and Evaluation**

In all these cases we have seen how, having identified a phenomenon X with the conditions that realize X in the causal structure of the world, the explanatory role that X plays within a theory is taken to be exclusively specified by citing X’s causal realizer. On account of this these theories are committed to D1 and D2. This is indeed a very stringent view of what naturalization requires with substantial reductionist implications about the theoretical role of a phenomenon. As in the case of D1, these implications are a consequence of the fact that D2 runs together two logically independent questions. The first is the *causal* question about the conditions that realize the theoretical role of a given phenomenon. This question is a demand to show that there are naturalistic conditions for giving the relevant explanation (or prediction). The second question, in turn, is the *constitutive* question about the theoretical role of a phenomenon. What makes these two questions logically independent is that the answer to the first one implies nothing, by itself, about the answer to the second one. So although a necessary condition for applying an explanation to a natural system requires that the conditions that realize that phenomenon be indicated, it doesn’t follow that such conditions enters into the content of the explanation. There are phenomena for which this inference holds, as per the case of the molecular genes. But no reason has been giving to think that all phenomena must conform to this inference. Hence the dogmatic character of the assumption. In the next chapter I will survey a number of examples in which the causal realization conditions for a phenomenon make no contribution to the specification of its theoretical role. In particular, when we are dealing with a phenomenon that has distinctive normative role to play in explanation, the eliminativist implications of D2 become problematic, as per the case of intentional action.

2.4. **Eliminativism and the Two Dogmas**

Eliminativism about teleological explanation is generated by the implicit commitment of the etiological and the mechanical approach to D1 and D2. Again, there is nothing wrong with eliminativism per se. It is the context of naturalizing normative phenomena such as teleological explanation that it becomes an acute problem.
2.4.1. Eliminativism and the First Dogma

As we have seen, D1 concerns the individuation of theoretical entities: To say what some phenomenon X is just is to say how X is caused or realized. The explicit aim of the mechanistic approach to teleological explanation surveyed in the preceding chapter is to give a naturalistic account of purposiveness. The methodological strategy of naturalization employed by this approach assumes that naturalization requires the specification of non-circular strictly causal-mechanical conditions for a system to be purposive. Insofar as this strategy gives a constitutive account of what it is to be a purposive system in terms of the conditions that realize purposiveness in the natural order, this strategy is implicitly committed to D1.

To say what purposiveness is according to mechanism, just is to say how it is realized. In each case, the strategy first identifies some condition that is necessary and/or sufficient for purposiveness to be realized in the causal-mechanical structure of the world. For example, according to cybernetics, the condition is a negative feedback mechanism. And according to the system-theoretic account the condition is plastic and persistent behavior. Finally, for the autonomous-systems view the condition is a mechanism of self-maintenance. Then, each account takes these conditions to be metaphysically constitutive of purposiveness and hence as providing a causal-mechanical principle of individuation. So according to cybernetics having a goal is nothing but being a feedback mechanism of certain regulatory kind, and according to the system-theoretical account having a goal just is for the system to behave in a certain way, and for the autonomous-systems account having a goal just is to possess a self-maintaining causal-mechanical organization. Purposiveness is thus not just located as part of the causal-mechanical structure of the world, it is identified with the mechanisms or behavior that causally realize them. So again, mechanism is committed to D1.

2.4.2. Eliminativism and the Second Dogma

To recall, D2 is about how theoretical entities figure in explanations: The theoretical role of a phenomenon X can be specified exclusively by citing the way X is causally realized. As we have just seen, each of the standard mechanistic accounts introduced in Chapter 1 identifies some putative causal-mechanical condition that is necessary and/or sufficient for purposiveness to be causally realized in the world. That’s an application of D1. On this basis mechanism derives an
account of teleological explanation. The idea is that we can specify the theoretical role that goals play in explanation in terms of the conditions that realize purposiveness in the causal-mechanical structure of the world. So mechanism is committed to D2.

For example, the explicit aim of the etiological approach is to give an account of teleological explanation. Etiology assumes that naturalizing teleological explanation requires the specification of non-circular strictly causal-mechanical conditions. Having identified having a function with its realizer—being a selected effect—as per D1, etiology then takes such conditions to specify exhaustively the role that functions play in explanation. On account of this etiology is committed to D2. According to the cybernetics account goals explain as part of negative feedback mechanism by sending signals to the system to adjust its trajectory from deviations. In turn, on the autonomous systems account goals explain by causally contributing to the self-maintenance of the system. The common denominator is the view that the theoretical role of goals in explanation can be exhaustively specified in terms of the realizer of that role, as per D2: since the conditions that are causally necessary for teleological explanation to apply in the world figure in the content of the explanation, teleological explanation is just an instance of some form of causal explanation. So teleological explanation is eliminated in favor of causal explanation.

2.4.3. How the Two Dogmas Generate Eliminativism

Eliminativism is the view that goals are part of the causal structure of the world but they don’t play a distinctive explanatory role as goals such that teleological explanation is just an instance of causal explanation. Mechanism and etiology assume that naturalizing purposiveness requires the specification of non-circular strictly causal-mechanical necessary and/or sufficient conditions for some entity to be purposive. So mechanism is committed to D1. Having identified purposiveness with its causal realizer, mechanism takes such conditions to exhaustively specify the role that goals play in explanation, thus dispensing with teleology all together. So mechanism is also committed to D2.

---

9 It may be that for a teleological explanation to apply to biological trait (token), the trait must be an instance of an adaptation (type). Having a causal-selective history may thus be a necessary causal condition for teleology to apply to a biological system. But it doesn’t follow that the historical explanation conveys the same information or has the same methodological role as the teleological one such that we can replace the latter with the former.
By identifying the realizer of purposiveness in the causal structure of the world mechanism satisfies the desideratum of naturalism. But importantly this in itself entails nothing about the individuation of goals or about the way the figure in explanations. So locating purposiveness in the causal-structure of the world does not, by itself, prevent us from assigning goals a distinctive explanatory role as goals. It is the assumption that the method of naturalization requires D1 and D2 that generates the eliminativist implications of mechanism. That is, it is the assumption that goals are individuated by their causal realizers (D1) and that goals explain as causal realizers (D2) that renders mechanism explanatorily inadequate as an account of teleological explanation. Hence eliminativism.

The problem with eliminativism about teleological explanation is that goals, unlike mere state of affairs, are characterized by their normative character as manifested in the distinctive normative role they play as goals in explanation. But, it seems, the causal realizers of goals are inapt to capture this normative character and hence this role. For example, unlike the case of the molecular genes, the main problem for CTA is that intentional explanations are characterized by their normative implications. These explanations are after all rationalizing: They explain by making rational sense of the behavior of the agent. The idea is that given an agent’s beliefs, desires and circumstances, the agent ought to act in a certain way. A good intentional explanation thus must convey the information that the agent was justified in acting in the way she did in light of her reasons and circumstances (Davidson 1963). So the explanatory role that intentional states play as reasons in action explanation is a distinctive normative role. But is not clear how brain states can capture this role qua brain states. It would seem that CTA faces the problem of elimination: by identifying reasons with their causal realizers reasons can’t explain as reasons, that is, they can’t provide rational justification. So although there is no problem in specifying the theoretical role that genes play in biological generalizations in terms of its causal realizers, there is a prima facie problem with specifying the theoretical role that intentional states play in rationalizing generalizations in terms of its causal realizers (more on this in Chapter 6).
2.5. Primitivism and The Two Dogmas

I will argue that primitivism is generated by the implicit rejection by the normative approach of D1 and D2. Again, there is nothing wrong with primitivism per se. It is in the context of naturalizing normative phenomena that it becomes an acute problem.

2.5.1. Primitivism and the First Dogma

The normative approach is motivated by an attempt to account for the distinctive character of teleological explanation. It starts with an account of the concept of a goal and then derives an account of the distinctive explanatory role that goals play as goals in teleological explanation on that basis. Goals, on this view, are irreducibly normative states of affairs. For example, according to the axiological account to say that $S$ did $x$ in order to $G$ is to say that that $S$ did $x$ because [$x$ contributes to $G$ and $G$ is good]. A state of affairs $G$ is a goal then only if $G$ is intrinsically good such that $G$ ought to be achieved by a purposive system $S$ (Bedau 1992). In turn, according to the rationality account to say that $S$ did $x$ in order to $G$ is to say that $S$ did $x$ because doing $x$ is the rational thing to do given $G$ and $S$’s circumstances (Sehon 2005). Goals are thus irreducibly normative features of the world. Hence, they cannot be specified in non-normative strictly causal-mechanical terms, so we cannot indicate the conditions that causally realized them and hence locate them in the causal-structure of the world. The normative approach thus rejects D1.

Thompson’s (2008, 10) neo-Aristotelian view in meta-ethics is an example of normativism rejecting D1. According to his approach, the analysis of the categories of life is an a priori affair that is independent of biology. So there is no need to say anything about how these categories are causally realized in the world. Thompson is quite vocal about the problem of primitivism:

“if anything is clear is that ‘the practical’ happens precisely in this so-called manifest image; the categories with which the reflecting and deliberating agent operates and conceives herself belong here [...] My hope, of course, is that my procedure will do something to defend the possibility of attaining truth here and there within this supposed image, and thus to oppose the globally fictionalist attack Sellars mounts”.

The idea is that at least some truths about the manifest image are primitive and hence cannot be located in the scientific image. Similarly Foot (2001, 22-23) talks about the primitive character of reasons against “those already in thrall to the mechanical picture of the workings of the mind?” She tells us,

“Why should we not take the recognition of reason for acting as bringing the series to a close? Recognition of a reason gives the rational person a goal; and this recognition is, according to the argument of the present chapter, based on facts and concepts, not on some prior attitude, feeling, or goal.”

By “prior attitude” she means the view that reasons need to be located outside the domain of reasons in the causal structure of the world. The idea is that since reasons give the rational person a goal they fulfill their explanatory role. So there is no need to further specify how this role is causally realized, hence primitivism.

2.5.2. Primitivism and the Second Dogma

D2 applied to goals is the view that, given that they are individuated by their causal realizers as per D1, goals figure in teleological explanations as realizers such that we can exhaustively specify their explanatory role by citing the conditions that causally realize them. As we have just seen, each of the standard normative accounts introduced in the Chapter 1 individuate goals in irreducibly normative terms and hence reject D1. On the basis of this normative account of goals normativism derives a normative account of teleological explanation. The idea is that we can specify the theoretical role that goals play in explanation only in irreducibly normative terms. So having rejected D1 normativism also rejects D2: we can’t exhaustively specify a goal’s explanatory role by citing its causal realizer. Rather, the theoretical role of goals must be specified in irreducibly normative terms, that is, either in terms of values as per the axiological account or in terms of reasons as per the rationality account. So normativism is silent about what the causal realizers of goals are, hence primitivism.

For example, the axiological analysis (Bedau 1992) holds that teleological explanations are irreducibly evaluative: $S$ did $x$ in order to $G$ iff $S$ did $x$ because [doing $x$ contribute to attaining $G$ and $G$ is good], where the ‘because’ has a wide scope such that the value condition itself—the goodness of the consequence—not just its consequences enter into the content of the explanation.
A teleological explanation, on this view, explains the occurrence of something in terms of its being good. In turn, according to the rationality analysis, teleology plays an irreducible rationalizing role in the interpretation of the behavior of rational agents (Sehon 2005). The idea is that to say that ‘S did x in order to G’ implies that, given G and S’s circumstances, S had reason or was justified in doing x or that S ought to have done x. These normative implications, furthermore, are constitutive of the content of the explanation. A teleological explanation, on this view, explains the occurrence of something as rationally justified. Since goals play an irreducibly normative role, teleological explanations have an irreducibly normative content. Even if we knew the conditions that causally realize goals in the causal structure of the world, we couldn’t specify their theoretical role by citing their causal realizers precisely because the specification of the theoretical role of goals requires the use of normative concepts. So primitivism rejects D2.

2.5.3. How the Two Dogmas Generate Primitivism

By assigning goals a distinctive normative role in teleological explanation normativism satisfies the desideratum of explanatory adequacy. But by remaining silent about the conditions that realize goals in the causal-structure of the world normativism fails to satisfy naturalism. Hence primitivism, the view that goals play a distinctive normative role as goals in explanation offers no account of goals as part of the natural order. Now, individuating goals in normative terms does not, in itself, entail that they can’t be naturalized, that is, that they cannot be located in the causal structure of the world unless we assume D1. In the same way, the view that the theoretical role of goals can only be specified in normative terms does not, in itself, entail that we can’t specify what the causal realizers of this roles are unless we assume D2. So normativism implicitly assumes that naturalization requires D1 and D2: locating goals in the natural order by showing how they are causally realized entails identifying them with their causal realizers and specifying their explanatory role by citing their causal realizers. Having assumed that naturalization requires D1 and D2, normativism rejects naturalization on the grounds that it fails to capture the normative nature of goals and the normative content of teleological explanations. But the problem with primitivism is that without locating goals in the causal-structure of the world, normativism makes a mystery out of the fact that teleological explanations applies to natural entities, such as biological organisms and psychological agents.
2.6. Conclusion

In this chapter I have offered a diagnosis of the dichotomy identified in the previous chapter between eliminativism and primitivism. I have argued that the dichotomy is predicated on what I’ve called two dogmas of naturalism: D1 is the view that to say what something is just is to say how it is caused or realized. And D2 is the view that the theoretical role of a phenomenon X can be specified exclusively by citing the way X is causally realized. After giving some examples of these dogmas in contemporary philosophy I have argued that while eliminativism is generated by the implicit commitment of etiology and mechanism to D1 and D2, primitivism is generated by the implicit rejection of D1 and D2 by the normative approach. These dogmas thus explain why etiology and mechanism satisfies naturalism at the expense of explanatory adequacy while normativism does the opposite. My goal has not been to refute these dogmas but to show how they motivate the stringent demands of the standard method of naturalization that generates the dichotomy. I have also suggested that by virtue of these dogmas the dichotomy generalizes to other cases involving normative concepts.

I believe each of the standard accounts gets some crucial aspect right about teleological explanation that I embrace for my project. What I reject is the set of assumptions that prevent us from having a unifying account that integrates the good in each. In particular, I believe that mechanism is right that an adequate account of teleological explanation should be naturalistically acceptable and that this requires locating goals in the causal structure of the world. Moreover, I think that each of the mechanistic accounts contributes to our understanding the conditions that realize causally purposiveness in the natural order. What I reject is the assumption, D1, that goals are individuated in terms of the conditions that causally realize them and that such condition allows us to exhaustively specify their explanatory role as per D2. So by abandoning D1 and D2 we can throw out the bath water of eliminativism without disposing of the naturalistic baby.

In turn, I believe that normativism is right that teleological explanation is a distinctive form of explanation and that its distinctive feature is partly the normative role that goals play as goals in explanation. I also agree with normativism that standard naturalism is unnecessarily stringent. But I disagree with normativism that we should therefore remain silent about the place of goals in the causal structure of the world. Normativism seems to assume that the standard strategy of naturalization as per D1 and D2 is the only available strategy. But in the next chapter I will argue
that this is not the case. By abandoning D1 and D2 we can throw the bath water of primitivism without rejecting normative character of teleological explanation and at the same time specify the conditions that causally realize goals in the world. Naturalism requires that we locate goals in the causal structure of the world by indicating the conditions that causally realize them. However, the standard method of naturalization takes such conditions to individuate X and specify the role that X plays in explanation. These are further theoretical commitments. The naturalist is free to appeal to this dogmatic strategy of naturalization. But naturalism per se does not necessitate it. So the naturalist is free to employ an alternative non-dogmatic hence less stringent strategy of naturalization in order to satisfy both, naturalism and explanatory adequacy.
Chapter 3
Three Grades of Naturalistic Involvement

3.1. Introduction

In the previous chapter I identified two assumptions or dogmas implicit in the standard strategy of naturalization. The first, D1, is about the individuation of theoretical entities: To say what some phenomenon \( x \) is just is to say how \( x \) is caused or realized. The second, D2, is about how theoretical entities figure in explanations: The theoretical role of some phenomenon \( x \) can be specified exclusively by citing the way \( x \) is causally realized. I argued that the dichotomy between eliminativism and primitivism identified in the first chapter is generated by the implicit commitment of mechanism and etiology to D1 and D2 on the one hand, and the implicit rejection of normativism of these dogmas on the other hand. In this chapter I explore what naturalism requires. The aim is to show that there is an alternative strategy of naturalization based in science that is not committed to these dogmas and hence provides a middle way between the two extremes of eliminativism and primitivism. This chapter thus concludes the philosophical analysis of the methodology of naturalization initiated in the first chapter.

Naturalism, recall, is the view that reality is exhausted by the immanent, self-contained space of causes that natural science theoretically describes. To be a naturalist about \( x \) is to believe that \( x \) has a place in this space. I will argue that there are a number of ways of ‘being natural’. To this effect I will outline criteria of naturalness—three grades—that vary in their stringency. According to the first and weakest grade (G1) introduced in section 3.2, for \( x \) to be natural is for \( x \) to be causally realized. This is just the basic commitment of naturalism to physicalism. The distinctive feature of G1 is that even if \( x \) has a place in the causal order of the world it need not have a role in the theoretical-scientific account of the world. In turn, the second grade (G2) introduced in section 3.3 is the view that for \( x \) to be natural is for \( x \) to be causally realized as per G1 and to play an indispensable role within a scientific theory irrespective of how that role is realized. The central insight behind G2 is that for a wide range of phenomena \( x \) familiar from scientific work on complex dynamical systems, \( x \)’s nature is specified by its scientific role, not by the realizer of this role. So G2 naturalism is not committed to D1 and D2. Finally, in section
3.4 I introduce the third and more demanding grade (G3): the view that for \( x \) to be natural is for \( x \) to be causally realized as per G1, have a distinctive theoretical role as per G2, and for that role to be specified in terms of its realizer. G3 adds an essentialist condition to G2 according to which \( x \)’s theoretical role and hence \( x \)’s nature is determined by its realizer. This is the standard strategy of naturalism as per D1 and D2. Unlike G1, G2 and G3 constitute legitimate forms of scientific naturalism. But while D1 and D2 entail G3 they don’t entail G2. I conclude in section 5 that D1 and D2 impose unnecessarily stringent demands on naturalism that risk the exclusion of a wide range of phenomena that are scientifically natural. G2 offers an alternative strategy that avoids the pitfalls of standard naturalism. In particular, G2 offers a scientifically legitimate alternative strategy of naturalization of teleological explanation that avoids the eliminativist and primitivist implications of the standard G3 strategy.

### 3.2. Grade 1: Causal Realization

According to the first grade of naturalistic involvement (G1), for a phenomenon to be natural is for it to have a place in a closed, complete, and connected space of causes, the causal order of the world. G1 expresses the basic metaphysical commitment of naturalism to physicalism. Accordingly, to naturalize a phenomenon is to point to its location in that space by showing how it is causally realized in the physical world (Jackson 1998). G1 offers the weakest criterion for being natural. It specifies a metaphysical condition for being part of the natural world. But it does not impose either a principle of individuation that determines the identity of the phenomenon or a methodological condition for being part of the scientific account of the natural world. So a phenomenon may be natural as per G1 even if it is scientifically irrelevant for predictive and explanatory purposes. They belong in the causal order of the world but they don’t figure as such in science. The motivation for G1 is typically to dispel the mystery about the natural ground of some putative non-natural phenomenon. Let us consider some examples.

#### 3.2.1. Transubstantiation

Consider the theological phenomenon of transubstantiation, the process by which the bread and wine used in the Christian ritual of mass are said to become the flesh and blood of Christ. There is no physical mechanism that can cause bread and wine to be transformed into flesh and blood, let alone into the flesh and blood of a particular person. Moreover, such a transformation is
excluded by scientifically established chemical principles. Therefore, transubstantiation cannot be causally realized in the physical world and hence is not a natural phenomenon. Rather, transubstantiation is a paradigmatic example of a supernatural phenomenon.

3.2.2. Demonic Possession

Now consider another theological phenomenon, demonic possession. Empirical research shows that most cases of demonic possession are caused by a certain kind of epilepsy in the temporal lobe or by some kind of psychotic disorder. Demonic possession is thus causally realized by neurophysiological structures in the brain. So according to G1, these experiences are natural as effects of these conditions because they can be located in the neurophysiological and psychiatric order of the world. So while transubstantiation lacks a place in the causal, physical order of the world, demonic possession can be located as a neurological or psychiatric effect. However, demonic possession plays no role within a scientific-theoretical account of the world as demonic possession. So even if the phenomenon of demonic possession has a place in the natural order the concept of demonic possession has no place in the scientific account of the natural order. But by explaining demonic possession as the effect of causal mechanisms the alleged supernatural character of the phenomenon is dispelled and hence the mystery removed. Although demonic possession is nothing metaphysically over and above a particular kind of epileptic episode, it doesn’t follow that a demonic possession just is the relevant neurophysiological structure causally responsible for it as per D1. There may be distinctive psychological and behavioral features of the episode that are constitutive of that particular kind of epileptic episode and that distinguish it from other kinds. So even if we can give a causal-mechanical account of demonic possession in terms of a particular kind of epilepsy we need not individuate that particular kind of epilepsy in terms of the neurophysiological mechanisms that cause or realize it, as opposed to, say, the behavioral profile of such an episode.\textsuperscript{10}

\textsuperscript{10} For an account of eliminativism about the mental see Churchland (1981).
3.3. Grade 2: Scientific Role

According to the second grade, for a phenomenon to be natural is for it to be causally realized as per G1 and to play an indispensable role in a scientific theory irrespective of how it is causally realized (see Tye 1992; Stich 1992; Burge 2010). G2 adds a methodological condition to G1. The rationale is that if \( y \) realizes \( x \) and everything that \( x \) predicts or/and explains can be predicted and explained exhaustively in terms of \( y \), then \( x \) is scientifically redundant and hence theoretically dispensable. So for \( x \) to earn its place in the scientific account of nature \( \textit{as } x \), its role must be non-redundant, that is, \( y \) need not figure in the specification of a \( x \)'s theoretical role. A phenomenon has an \textit{indispensable} role in a scientific theory if it is necessary to predict, explain or classify some further phenomena. G2 also adds a metaphysical condition to G1. In the second chapter the principle of \textit{theoretical individuation} (TI) was introduced according to which, to say what something is just is to say what its theoretical role is. That is, the theoretical role of some phenomenon \( x \) is exhausted by its individuation conditions. Accordingly, the central idea behind G2 is that, for a wide range of phenomena, its nature is specified by its scientific role, not by the causal realizer of this role. Let us consider some examples.

3.3.1. Thermodynamics and Statistical Mechanics

Thermodynamics is a phenomenological theory that aims to understand the bulk or gross macroscopic behavior of systems that exchange work and heat with the conditions in which they are embedded. To do so the theory appeals to a set of macroscopic properties or quantities including temperature (\( T \)), energy (\( U \)), entropy (\( S \)), volume (\( V \)) and pressure (\( P \)). These physical quantities are defined in terms of the macroscopic laws that govern how thermal behavior changes with time under various circumstances as they tend towards equilibrium. In particular, the Zeroth law defines temperature, the first law defines energy, and the second law defines entropy (Atkins 2010). Less general laws relevant for thermodynamics include the ideal gas laws or Boyle-Charles law. These laws specify the role that these quantities play in the prediction and explanation of the gross macroscopic dynamics of these systems irrespective of the physical conditions that causally realize it at the microscopic level.

For example, the macroscopic state of a homogenous gas in a container is governed by the ideal gas law \( PV = kT \) (\( k \) is a constant) which states that for a specific quantity of gas the product of \( V \)
and $P$ is proportional to $T$. Information about these macroscopic parameters together with the macroscopic law is sufficient to predict and explain the system’s later macro-state. The microscopic conditions that causally realize these macroscopic properties and hence make the application of these parameters and laws physically possible does not figure in their content. Hence, the realizers of $P$, $V$ and $T$ do not figure in the specification of their role within the theory of thermodynamics. The role of these concepts is to capture the gross macroscopic disposition of the whole system to behave in specific ways in response to specific macroscopic conditions irrespective of how this behavior is causally realized.

Statistical mechanics, in turn, aims to understand thermodynamic behaviour in terms of the microscopic mechanical laws that govern the motion of the particles that causally realize it. Statistical mechanics thus locates thermodynamic behavior in the causal order of the world as per G1. To do so, the distinctive methodological strategy of statistical mechanics is the use of probabilistic assumptions to construct an idealized model of the behaviour of individual particles. In particular, by means of averaging techniques the model abstracts away the mechanical details about the properties of the individual particles that realize thermodynamic behavior. This allows the model to appeal to statistical aggregates or ensembles rather than the computationally intractable individual molecules.

This kinetic theory of gases, for example, explains changes in the temperature of a gas by citing the average kinetic energy of the molecules that causally realize it, irrespective of the properties of the individual molecules that causally realize it. A model of the temperature of a gas assigns a probability distribution over the constituents of the system by means of an ensemble of replicas of the system or a ‘universal class’. The thermal state of the gas is explained by deducing its actual state from the ensemble. This distribution is assumed to be unaffected by a large proportion of microscopic facts about individual interactions among those constituents. For example, it is assumed that the behavior of each molecule is stochastically independent from the behavior of other molecules. It is also assumed that, except during the elastic collisions between molecules and the walls of the container, no forces act on the molecules. Instead, the distribution is determined only by macro-level facts about the whole system, such as the total number of molecules (Strevens 2005).
Certain microscopic causal details of the properties of individual particles thus make no difference to the mechanical derivation of macroscopic thermodynamic behavior. The statistical mechanical concept of temperature is defined over ensembles or populations of molecules. Since there is no set of individual molecular properties the possession of which is necessary and sufficient for the concept of temperature to apply to a system, there is no constitutive analysis or definition of temperature in exclusive microscopic terms. So the role of the concept of temperature in the prediction and explanation of the global dynamics of thermodynamic systems cannot be exhaustively specified in terms of the microscopic conditions that causally realize that role. Temperature is a gross dispositional property of the system as a whole that can be specified in terms of other macroscopic concepts used in the explanation of the global dynamics of thermodynamics systems.

Temperature is realized by the mean molecular kinetic energy only in gases. It is realized by the black body distribution of electromagnetic waves in a vacuum and in plasma there is temperature even if there are no constituent molecules. A gas made up of molecules and radiation made of wavelengths of light can have the same temperature. Temperature can hence be universally instantiated across systems with very different physical constitutions. This universality enables us to explain and predict when systems will be in thermal equilibrium regardless of their composition and size.

Furthermore, due to its universality, thermodynamic behavior is stable across different molecular configurations of the system (Batterman, 2000; Kadanoff, 2000; Strevens, 2005). The particular microscopic configuration is thus sufficient but not necessary to causally realize the macroscopic state. So the macroscopic state supervenes on its microscopic realizer. This modal fact about the behavior of the system is a macroscopic fact that characterizes the distinctive profile of thermodynamic behavior. The macroscopic behavior of the system thus contains indispensable modal information about the system that is absent at the microscopic level. For example, the temperature of a system tells us the most probable distribution of populations of molecules over the available state of a system at equilibrium (Atkins, 2010). If the temperature is increased, the

11 This indicates that there is no one-to-one map between thermodynamic concepts and statistical mechanical concepts. Rather, a given thermodynamic concept corresponds to many different statistical mechanical ones (Sklar, 2009). At most there is an identity between ensemble quantities and thermodynamic ones.
population will go to higher energy states and if it is decrease the population goes back to the states of lower energy. In a slogan, more is *modally* different. Since macroscopic behavior exhibits a distinctive modal profile—it supports a different set of counterfactuals from those of the microscopic structure that realizes it—it calls for distinctive set of concepts.

Thermodynamic behavior exhibits maximal insensitivity to microscopic composition at critical points at phase transition (see Batterman 2000; 2005; Rueger 2000; Wayne and Arciszewski 2009; Morrison, 2012). A phase transition occurs when a physical system undergoes a dramatic change in its global properties. For example, water turns into vapor when it is heated to its boiling point or a solid transforms into a liquid when ice melts or a magnet loses its magnetism when heated. Phase transitions are characterized by an order parameter, such as the density or the magnetization, which changes as a function of a property of the system, such as temperature. The special value of the parameter at which the system undergoes a phase transition is the system’s critical point. The closer the parameter is to its critical value, the less sensitive is the order parameter on the details of the system. Systems with very different physical constitutions have the same critical exponents and hence exhibit the same kind of universal behavior at phase transition. To derive thermodynamic phenomena at phase transition from the statistical mechanics of its underlying molecular structure, the theory requires an infinite idealization according to which the number of particles is infinite (i.e. \( N = \infty \)) despite the fact that the system is actually composed of a finite number of particles.\(^\text{12}\) As Kadanoff (2000, 238) puts it,

> “The existence of a phase transition requires an infinite system. No phase transitions occur in system with a finite number of degrees of freedom.”

Thus, the statistical mechanical theory represents the thermodynamic phase transition as a mathematical singularity. The infinite idealization of the thermodynamic limit is thus indispensable for statistical mechanics to account for phase transitions.

> “A major triumph of modern statistical physics was its theory identifying the universal aspects of critical phenomena—a theory in which the macroscopic

\(^{12}\) ‘\(N\)’ is the partition function for the components of the system. The idea is that a finite \(N\) cannot exhibit the nonanalytic behavior required to represent qualitatively different states of matter (Kadanoff 2000).
(thermodynamic) properties of a system near a phase transition are insensitive to the particularities of the system, namely its underlying microscopic properties.” (Fox-Keller 2009, 22)

In conclusion, thermodynamic properties and the system that instantiate them are causally realized by their underlying physical constituents. By means of probabilistic idealizations over the properties of individual constituents and interactions we can derive and hence explain how macroscopic behavior emerges out of its microscopic realizers. However, since these idealizations imply that thermodynamic regularities are insensitive to microscopic details, the very same methodological strategy prevents the identification of thermodynamic quantities with unique microscopic structures. Temperature is natural, in this sense, because it is causally realized and has a place in a comprehensive scientific theory irrespective of its causal realizer.

3.3.2. Fluid Dynamics

Less general than thermodynamics, fluid dynamics is the study of the bulk or gross macroscopic behavior of fluids, such as liquids and gases. A fluid is defined as any substance that deforms continuously when subject to a shear stress (McDonough 2009). For example, applying a shear stress to a sample of water, however small, produces a deformation such that the more stress is applied the more the water will deform. In contrast, applying a shear stress on a solid block of steel will produce no change or deformation until an extreme amount of stress has been applied. A pile of sugar does not deform, even if it flows, so it is not a fluid. A fluid is characterized by a set of dispositions to behave in certain ways in response to certain stimulus conditions. Some of the most prominent modes of behaviors that a fluid exhibits are turbulence, viscosity, mass diffusivity and thermal conductivity. These macroscopic properties enable us to distinguish one fluid from another and to predict and explain its behavior under various conditions. Fluids are of course composed of billions of randomly moving molecules that collide with one another. These molecular conditions causally realize fluid behavior. So fluids are natural as per G1.

The behavior of fluids is governed by a set of non-linear partial differential equations known as the ‘Navier-Stokes equations’. They describe a fluid’s behavior under stress by calculating various thermodynamic variables such as, pressure, density, and temperature. Despite the fact that the material constitution of fluids consist of discrete entities (particles), to model their
macroscopic behavior with these equations fluids are assumed to be a continuum and hence infinitely divisible. This continuum idealization enables us to define properties like density, temperature, pressure and velocity at infinitesimal small points that vary continuously from one point to another. This means that we can associate within any volume of fluid, no matter how small, those macroscopic properties that we associate with the bulk fluid (McDonough 2009). Similar to statistical mechanics, the continuum assumption is achieved by averaging the velocities of all molecules in a region surrounding a given point. On account of the continuum assumption the scientific understanding of the bulk behavior of fluids is largely insensitive to the conditions that causally realize them. In the words of Goldenfeld and Kadanoff (1999, 87-88),

“In fluid dynamics the large-scale structure is independent of detailed description of the motion of the small scales. [...] To get these gross features, one should most often use a more phenomenological and aggregated description, aimed specifically at the higher level [...] by trying to separate universal scaling features from specific features. The inclusion of too many processes and parameters will obscure the desired qualitative understanding.”

Fluids are thus defined macroscopically in terms of the set of bulk behaviors that characterize them, not in terms of the mechanical laws that govern the behavior of the microscopic constituents that causally realize them. So the molecular conditions that causally realize a fluid do not figure in its individuation conditions. In most cases we cannot tell whether two systems are the same fluid just by looking at the behavior of the individual molecules that compose it. Rather, fluids are individuated in terms of their macroscopic patterns of deformation behavior in response to shear stress conditions. Even if certain molecular conditions must be in place for the concept of fluid to apply, there is no categorical set of molecules the number of which is necessary and sufficient for the concept to apply. So the role that the concept of a fluid plays in fluid dynamics cannot therefore by exhaustively specified by citing its causal realizers. In fact, for the definition of a fluid to be consistent with the continuum hypothesis the content of the concept must be exhausted by those macroscopic patterns of deformation behavior. The concept is characteristically vague from a micro-structural point of view. Yet, there are paradigmatic

---
13 Part of the motivation for the continuum assumption is that, due to random motion, at any giving point and an any given instant there will be no molecules and hence no way to define the velocity of the fluid at that point.
macroscopic patterns of deformation behavior in response to paradigmatic shear stress conditions that provide prototypical conditions of application. Fluids are thus natural because they are causally realized as per G1 and because they play a distinctive theoretical as such irrespective of their realizers as per G2.

Now consider a distinctive property of fluids as fluids, viscosity. Viscosity is the disposition of a fluid to resist shear stress (Massey & Ward-Smith 2011). Water deforms more readily than honey when the same shear stress applied to both because the higher viscosity of honey offers more resistance. The viscosity of a fluid is causally realized by the friction between the particles that compose it moving at different velocities as a result of their shape, size and attractions. So viscosity is natural as per G1. But even if being thus composed is a causally necessary and sufficient condition for a fluid to have a certain viscosity, it doesn’t follow that viscosity is the concept of being composed of a specific configuration of individual molecules. So we can’t say what viscosity is just by citing how viscosity is causally realized. As the continuum assumption indicates, viscosity is a macro-level concept whose role in fluid dynamics is to capture macro-level facts about the complex dynamics of the whole ensemble of molecules irrespective of how this macroscopic dynamics is microscopically realized (Strevens 2005). A fluid’s viscosity specifies the way it behaves in certain contexts, that is, when shear forces act on it. So the conditions that causally realize viscosity do not figure in the specification of its role within fluid dynamics. For a fluid to have a particular viscosity then is for it to have a particular dynamics or behavioral profile such that it will respond in certain ways under certain conditions. So the molecular conditions that cause or realize or compose viscosity do not figure in its individuation conditions and hence in the specification of its role. So viscosity is natural as per G2.

The property of having a certain viscosity is individuated in terms of the macroscopic patterns of resistance behavior to deformation under shear stress. Hence, there is no set of molecules the number of which is necessary and sufficient for the concept of viscosity to apply and therefore no way to specify the role that the concept plays within fluid dynamics in terms of molecular conditions that realize it. The content of the concept is rather specified in terms of the macroscopic patterns of resistance behavior to deformation under shear stress. Viscosity cannot be identified with molecular dynamics on the grounds that the idealization of a fluid as a continuum, despite the discrete character of its underlying molecular dynamics, plays an indispensable role in the explanation of resistance to shear stress behavior.
3.3.3. Population Genetics

Organisms have traits that affect their individual fitness and hence their chances of surviving and reproducing. The activities of organisms with those traits produce changes in the structure of trait types in the population they compose. These changes happen when there is an increase or decrease of the relative frequency of heritable trait types in the population. To predict and explain the rate of change in their relative frequency, trait types are assigned fitness values that generate a probability distribution around the possible changes of a trait’s relative frequency. Evolutionary changes in the trait structure of a population are predicted and explained by various factors such as migration, mutation, natural selection and genetic drift. Selection occurs when there is variation in trait fitness. The expected outcome of this variation is some determinate change in the trait structure of the population. In turn, drift occurs when the change in frequency varies from the expected outcome due to random sampling. So selection represents the expected outcome while drift represents the deviation from the expected outcome. The role of these concepts is to predict and explain the large-scale statistical regularities that hold at the level of the whole population. These population-level regularities are causally realized by the activities of or life histories of individual organisms. In particular, in the case of selection, variation in trait fitness is realized by variation in individual fitness. So trait fitness is natural as per G1.

But according to a recent interpretation of evolutionary theory—the statistical interpretation—there is no constitutive definition or analysis of trait fitness in terms of individual fitness.\(^\text{14}\) In other words, variation in individual fitness is neither necessary nor sufficient for variation in trait fitness. Selection and drift are defined in terms of the fitness of trait types (Ariew & Lewontin 2004; Walsh, Lewens & Ariew 2002). The fitness of types is not determined just by the properties of the individual organisms of the population that posses tokens of that trait but also by statistical parameters that pertain to populations, such as demographic factors. Trait types are idealizations that abstract from the causal details of actual individual interactions by means of probabilistic assumptions that allow these statistical models to be projected to actual and counterfactual populations (Walsh 2013c). Selection and drift are thus statistical parameters and

\(^{14}\) This is not the standard interpretation of evolutionary theory. But my aim is not to defend the statistical interpretation or to assume it as the correct interpretation. My aim is to show that under a plausible interpretation of population genetics it constitutes an instance of G2.
the processes of selection and drift are stochastic processes. Due to the abstract, idealized character these statistical models of population dynamics, the large-scale statistical regularities that hold at the level of the whole population are largely insensitive to the causal ecological details that hold at the level of the individuals that compose the population. On account of this there is no constitutive definition of these statistical concepts in strictly causal-mechanical terms. Trait fitness is a primitive concept that plays an irreducible statistical role in population genetics.

This insensitivity to causal details reflects the fact that the large-scale statistical regularities that hold at the level of the whole population are robust across a wide range of changes in the properties of the individuals that compose them. That is, the values assigned to trait fitness are robust across perturbations in the individual fitness that causally realized them. This robustness is manifested in the fact that population-level regularities support a distinctive set of counterfactuals according to which, the distribution of trait types in the population would have occurred regardless of various changes in the details of the life histories of individuals of that population. Thus, the large-scale statistical regularities have a distinctive modal profile from that of the individual-scale causal regularities that realize them and hence contain indispensable modal information about the evolutionary dynamics of the population. Hence, changes in the trait structure of a biological population are explained and predicted in terms of the average fitness of the trait types that compose the population—a higher order effect—not in terms of the causal properties of the individual tokens that realize them.

“The modern genetical theory of evolution explains without tracing the causal processes and interactions represented by the multitude of individual life histories (life, death and reproductive events). Instead, modern evolutionary theory explains by abstracting radically from the details of the causal processes involving particular individual life histories. [...] Evolutionary explanations of this kind focus on identifying higher-level statistical variables that aggregate over many individual causal processes and that figure in general statistical regularities (or laws) that govern evolutionary events.” (Ariew et al. 2014, 16)

The ecological interactions among individual organisms partly explains how the distribution of trait fitness arises, thus providing the causal-mechanical conditions for the statistical models of populations dynamics to apply. But such causal-mechanical conditions do not figure in the
content of those models and hence in the explanations and predictions they elicit. Trait fitness supervenes not just on individual properties but also in other population-level properties. These models tell us why the population tends to change in regular reliable way, not how the fitness of trait types arise from individual fitnesses. There is then no definition of trait fitness in terms of the properties of individual organisms (variation on individual fitness) on the grounds that, for a given distribution of fitness, there is no arrangement of the causal properties of organisms that is necessary and sufficient to specify the values of trait types (Ariew and Lewontin 2004; Earnshaw-Whyte 2012; 2-13). So fitness distributions have a causal basis on individual fitness but it has no individuation in those properties. We cannot say whether two trait types have the same fitness just by looking at the individual fitness of tokens of that trait. Individual-level facts provide no demarcation criterion. The concept is defined in irreducible statistical terms by the role it plays in the explanation of population dynamics. So population-level facts are vague from the perspective of individual-level facts. Yet, paradigmatic instances of populations undergoing selection provide appropriate or prototypical conditions for the concepts of population genetics to correctly apply, admitting of course borderline cases, and hence for dynamical explanations of emergent behavior to hold.

3.3.4. Lessons from G2: Naturalism and Emergence

We have seen how emergent phenomena studied by the sciences that deal with the dynamics of complex systems, from condensed matter physics to population biology, provide paradigmatic examples of G2. Complex systems are composed of many microscopic parts that strongly or weakly interact in linear or non-linear ways from which a global macroscopic pattern of behavior or order emerges. But there is no unique microscopic structure or model that is consistent with the macroscopic behavior of the system. Rather, there is an equivalent class of microscopic models all of which exhibit the same macroscopic behavior (Goldenfeld et al., 1989; Goldenfeld & Kadanoff, 1999; Kauffman, 1993). This behavior is ‘universal’ in the sense that systems with

---

15 Foundational work in complex systems dynamics includes Anderson (1972); Bak et al. (1987); Gell-Mann (1994); Haken (1983); Holland (1998); Kauffman (1993); Mandelbrot (1997); Prigogine & Nicolis (1977); Simon (1962); Turing (1952).

16 I take the micro-macro distinction to be relative to a system and to a given scale of organization. So for example, relative to a fluid the molecular dynamics is the micro level while relative to a population undergoing selection the micro level is the set of individual organisms or traits.
very different microscopic material constitutions can instantiate the same macroscopic behavior. The emergent order of the system is thus stable across a range of microscopic conditions (Batterman 2000; 2005; Kadanoff 2000; Strevens 2005). Consequently, although the macroscopic behavior of these systems is realized by its material constitution as per G1, it is largely insensitive to the details of their causal realization. This insensitivity allows us to develop theories about emergent phenomena that are largely independent of theories about their causal realization. So against D1, emergent theoretical entities are not individuated in terms of their causal realizers but in terms of the set of behaviors that characterize their global dynamics. And against D2, the role that emergent phenomena play within their respective scientific theories is not exhaustively specified by citing the conditions that causally realize them. As Thom (1972, 78) put it,

“Knowledge of the fine structure, molecules for a fluid, cells for an animal, is practically irrelevant for understating the global structure ... of the total system”.

Although being a gas or a fluid as well as temperature, viscosity and trait-fitness are realized by the underlying causal-mechanical properties of the individuals that compose them, these conditions do not figure in the specification of their theoretical role and hence their individuation conditions. The macroscopic behavior or gross dynamics of a complex system may be described either in a fine-grained way in terms of the local or intrinsic properties of its component parts, such as the position and momentum of individual molecules. It may also be described in a coarse-grained way in terms of the global or bulk properties of the system as whole, such as its temperature or entropy. But pace the standard strategy of naturalization, we cannot dispense with the coarse-grain description without losing crucial theoretical information. According to prominent account of explanation in the philosophy of science, an explanation should tell us how things could have been otherwise than they actually are. That is, how sensitive is the outcome or explanandum to changes in the conditions or explanans (Garfinkel 1981; Walsh 2012; Woodward 2003). The robustness of universal behavior implies that had microscopic conditions been different across a certain range of variations, the macroscopic state would still have been achieved. So knowing the particular microscopic conditions that produced the macroscopic state on a given occasion fails to convey the information that the macro state is reliably and robustly achieved. The particular microscopic structure that causally realizes an instance of macroscopic behavior on a given occasion is thus causally sufficient but not necessary. So universal behavior
exhibits a distinctive modal profile from that of the microscopic conditions that realize it, and hence contains exclusive modal information that is indispensable to understand how the system as a whole behaves.

According to the indispensability argument for the existence of a given domain of entities, we ought to be ontologically committed to all and only those entities that are indispensable to our best scientific theories (Putnam 1979; Quine 1980; Colyvan 2001). Emergent entities are indispensable to our best scientific theories about the gross behavior of complex systems, such as thermodynamics, statistical mechanics and fluid dynamics. Therefore, we ought to be committed to these entities as such. We cannot eliminate them in favor of the molecular configurations that realize them, as they are constitutive of a distinctive set of macroscopic empirically observable and counterfactual-supporting regularities. Emergent phenomena are thus natural but not ‘really something else’.

3.4. Grade 3: Scientific Essentialism

According to the third grade of naturalism (G3), the most demanding of all, to be natural is to be causally realized as per G1, to play an indispensable role in a scientific theory as per G2 and for that role to be completely specified by its causal realizer. The causal realizer of a phenomenon not only locates the phenomenon in the causal order of the world as in G1 but also provides the necessary and sufficient conditions that individuate that phenomenon and specify its role within a scientific theory. These conditions constitute the essential properties that ground similarity relations between members of the same natural kind. Since these conditions are discovered a posteriori by empirical methods, the essences they specified are scientific. We can call this grade essentialism (see Bigelow, Ellis, Lierse 1992; Bird 2007; Ellis 2001; Kripke 1980; Putnam 1975).

On account of its essentialist character, G3 has a characteristic logical form and metaphysical interpretation. It consists of a biconditional statement of the form ‘necessarily \( x \) is \( G \) if and only if \( x \) is \( M' \) where ‘\( M' \)’, the realizer of ‘\( G \)’, is a set of lower-level intrinsic properties and ‘\( G \)’ some higher-level property. This biconditional is meant to be a constitutive metaphysical account, not just a necessary equivalence. So it must be interpreted as expressing a metaphysical asymmetry between consequent and antecedent that expresses the ontological dependence of the former on
the latter according to which the consequent specifies the essence or principle of individuation of
the antecedent. The essence of something is the set of necessary and sufficient conditions that
makes it a member of its kind on the basis of which its behavior can be explained. Things of the
same kind share certain intrinsic properties and structure that explain their manifest similarities.
Individuation, in turn, is the principle of ontological determination that makes an entity one
entity, distinct from others and the very entity that it is as opposed to any other (Lowe 2009). 17
So G3 offers real definition, a criterion of demarcation. What makes this real definition into a
naturalistic account or naturalization is the fact that the consequent must be stated in non-circular
strict causal-mechanical terms of the kind endorsed by natural science. Let us consider some
examples.

3.4.1. Dispositional Essentialism

According to dispositional essentialism, an increasingly influential anti-Humean metaphysical
interpretation of science, fundamental natural properties are individuated by their causal powers
or dispositions such that the essence of a property and hence its identity is determined by its
potential causes and effects (Ellis and Lierse 1994; Ellis 2002; Chakravartty 2007; Bird 2007).
Natural kinds have all of their powers, capacities and propensities as a matter of metaphysical
necessity and this can be determined by scientific methods. Natural fundamental properties are
thus essentially distinguished from each other by what they dispose their bearers to do
(Shoemaker 1980). The idea is that certain essential properties of objects not only determine the
nature of these objects but also how they will behave in any situation. As Ellis (2002, 3) puts it

“Things behave as they do … not because of any external constraints that force
them to, but because this is how they are intrinsically disposed to behave in the
circumstances.”

The idea is that at least the fundamental laws of nature are grounded in the essential dispositional
properties of the things they are said to describe and hence are metaphysically dependent on
them. Laws are thus immanent in the world rather than being imposed by the forces of nature (or

17 If it is a property or relation rather than an individual, the principle of individuation is what determined which
sequence of objects (in any possible world) are instances of that property of relation.
3.4.2. Chemical Microstructural Essentialism

Natural essentialism is not limited to fundamental natural properties however. One prominent version of essentialism in the philosophy of chemistry is micro-structuralism, the metaphysical view that chemical kinds such as elements; compounds, etc. are individuated solely in terms of their microstructural atomic properties. In short, microstructure fixes the identities of chemical substances. This view is essentialist in that there are internal, microstructural properties of things by which they can be categorically divided into kinds from which their behaviour can be explained. The essentialist principle states that, necessarily, a sample of a chemical substance, X, is of the same kind as another, Y, if and only if X and Y have the same microstructure (Kripke 1980; Putnam 1975). Scientific essentialism about natural kinds implies that chemical substances have their molecular structures essentially and that chemical elements have their atomic number essentially.

Consider chemical elements. Chemical elements familiar from the periodic table are the building blocks of chemistry. They are pure chemical substances that consist of a unique kind of atom that is distinguished by the number of protons in its nucleus, which in turn determines nuclear charge. What makes nuclear charge rather than nuclear mass apt to determined the extension of the concept of an element is its capacity or disposition to survive chemical change and hence to gain and lose electrons. Chemists are interested in understanding chemical change and nuclear charge provides the most invariant property across chemical changes. Hence, it

---

18 The essence of these properties may or may not be taken to be dispositional. So although dispositional essentialism is compatible with chemical micro-structuralism it is not entailed by it.

19 According to recent developments in the philosophy of chemistry, when we move from chemical elements to chemical compounds where the bonding structure becomes relevant, micro-structuralism becomes problematic (Brakel 2000; Hendry 2008; Needham 2000). I will only addresses chemical elements.

20 The case of isotopes, which can have the same nuclear charge but differ in nuclear mass indicates the primacy of atomic number.
explains both the direction of chemical change and the properties of the substances they compose (Hendry 2008). The atomic number thus explains the chemical behavior of the element and hence specifies its classificatory, predictive and explanatory role in chemical theory. So micro-structuralism about chemical elements is the view that chemical elements have atomic essences determined by their atomic number such that the atomic number is sufficient to individuate the element and explain its role within chemistry. What makes some chemical phenomenon into the phenomenon that it is just is its underlying atomic structure. On account of this chemical elements are natural be virtue of having a causal essence as per G3.

For example, for something to be oxygen it is necessary and sufficient to have a certain atomic structure, namely, eight protons. This atomic structure not only locates oxygen in the causal order of the world as per G1, it also individuates instances of oxygen as oxygen thus providing a criterion of demarcation that categorically distinguishes instances of oxygen from any other chemical element. Furthermore, not only can the behavior of oxygen be explained by reference to its atomic structure, the role that oxygen plays in the explanation of further chemical element can be exhaustively specified in terms the atomic structure that realizes it. Thus, showing how some chemical element is causally realized provides a constitutive analysis of what the chemical phenomenon is and what role it plays is science. To state what some chemical phenomenon is just is to specify its underlying atomic structure. So knowing the atomic conditions that realize a chemical element gives an account of what the kind is and exhausts its theoretical role.

3.4.3. Developmental Microstructural Essentialism

We have seen scientific essentialism applied to fundamental physics and chemistry. Let us now consider a much more controversial case of essentialism applied to biology. According to Rosenberg (1997) the vast diversity of biological forms can be explained in terms of a small number of molecules including DNA, RNA and proteins and their interactions. Developmental biology is thus reducible to molecular biology. Given a total molecular description of the fertilized egg—the total DNA sequence and the location of all proteins and RNA—one could predict how the embryo will develop (Wolpert 1994). A strictly molecular-level description of the embryo and the environment at conception including the properties of DNA, RNA and proteins (and no reference to functional kinds) thus provides necessary and sufficient condition for explaining how all the molecules will be arranged in later times. No higher-order non-
molecular constraint such as epigenetic factors figures in the derivation of the adult embryo. Functional kinds in biology are thus not only causally realized in the biochemical structure of the world, which in turn is realized in the physicochemical structure of the world. They are also individuated by their biochemical structure, which specifies their role in developmental biology.

Rosenberg explicitly denies what he calls “the principle of autonomous reality” according to which phenomena, properties or kinds identified in functional biology are real and irreducible in the sense of “reflecting the existence of objective explanatory generalizations that are autonomous from those of molecular biology” (446). He claims that although initially the relevant molecular processes can “only be individuated in terms of the phenomena the factors are invoked to explain”, namely, the development of biological form, ultimately these processes are to be individuated in a way that is “independent of the phenomena they are invoked to explain” (456). Causal realization is thus individuation. Rosenberg also denies what he calls “the principle of explanatory primacy”, which tells us that functional-level explanations are “at least sometimes” indispensable for understanding underlying molecular processes. This principle entails that the molecular conditions that realize biological forms completely specify the theoretical role of these forms such that citing functional kinds for explanatory or predictive purposes is ultimately dispensable. Together these principles give us what we might call “molecular essentialism” about developmental biology.

“Cellular structures only come into existence through the molecular processes that precede them. There is in developmental molecular biology therefore no scope for claims about the indispensable role of cellular structures in these molecular processes” (455).

On account of these set of commitments Rosenberg’s account of development constitutes an instance of G3.

---

21 By ‘functional’ biology Rosenberg (446) means “non-molecular biology”. Functional biology “identifies biological structure and system by their causal roles, usually their adaptationally selected effects”.
3.5. Conclusion: Naturalism without Dogmas

We are now in a position to assemble the lessons from our discussion so far. First, there are a variety of ways that something could legitimately be said to be natural. G1 is the weakest. Saying that x is G1 doesn't guarantee it any indispensable role in a natural science. For that x must be natural in the sense of either G2 or G3. In the discussion of the problem of natural teleological explanation from the last chapter we said that eliminativism and primitivisms alike apply D1 and D2. But D1 and D2 together entail G3, essentialism, but not G2, emergentism: The two dogmas of naturalism are taken to say that x is natural only if x has a causal dispositional essence. This is a really stringent set of demands on naturalism. Arguably most of the phenomena of the natural sciences don't meet them. These dogmas seem to suggest a very rarefied conception of naturalism, one that is generally unsuited to the natural sciences. It risks implying that temperature, pressure, viscosity, entropy, tensile strength, compressibility, liquid, solid, gas, force, fitness and the like are not natural phenomena.

The lesson from G2 is that due to its universality, the macroscopic behavior of complex systems is largely insensitive to the microscopic details that causally realize them. Hence, microscopic details are mostly irrelevant to understand the macroscopic behavior of the system. Macroscopic concepts relate macroscopic facts to macroscopic facts and these relations exhaust their content. The conditions that cause or realize these phenomena are not part of their individuation conditions. On account of this emergent phenomena violate D1 and D2. Yet, these phenomena are perfectly natural. G2 thus suggests an alternative strategy of scientific naturalization. Instead of the single stage of giving a causal-mechanical definition, naturalization proceeds in two different stages. The first stage consists in locating x in the causal order of the world by showing how it is causally realized as per G1. The second stage, in turn, consists of giving a constitutive account of what conditions individuate the phenomenon in question as the phenomenon that it is in terms of its theoretical role as per G2, that is, in terms of the role that it plays in the explanation and prediction of the overall dynamics of a complex system irrespective of the realizer of that role but rather in terms of the set of behaviors that characterize it.

G3 is a legitimate strategy of scientific naturalization. But as the examples of G3 indicate, it doesn’t apply to all natural phenomena. In particular it fails for emergent phenomena. I take it that purposiveness is an emergent phenomenon characteristic of the behavior of highly complex
adaptive systems such as organisms. So as an alternative strategy of scientific naturalization, G2 has important implications for the project of giving an account of teleological explanation that is both naturalistically acceptable and explanatorily adequate. Against D1, G2 allows us to specify the conditions that causally realize goals in the natural order as per G1 without thereby identifying goals with their causal realizers. And against D2, the theoretical role that a goal plays in teleological explanation need not be specified by citing its causal realizer but by citing other conditions at the same level as per G2. So G2 allows us to occupy a middle way between the extremes of eliminativism and primitivism and hence to both, locate goals in the causal structure of the world and hence satisfy naturalism, and to assign them a distinctive role as goals in explanation, thus satisfying explanatory adequacy.
Part II
Agency as an Ecological Concept
Chapter 4
An Ecological Theory of Natural Agency

4.1. Introduction

The first three chapters of the dissertation have been about the methodology of naturalization. In the previous chapter (Chapter three) I proposed an alternative strategy of naturalization G2: To naturalize a phenomenon \( x \) consists of two stages: (i) show how \( x \) is causally realized in the world, and (ii), show that \( x \) plays an indispensable role as \( x \) in a scientific theory irrespective of how that role is causally realized. In this chapter I apply G2 to the phenomenon of agency and teleological explanation to give a novel naturalization.\(^{22}\) I argue that while both accounts share the same account of the place of goals in the causal structure of the world as per (i), the concept of a goal plays a different but related role in a theory of the nature of agency and in a theory of teleological explanation as per (ii). The theory of agency I propose is an ecological theory. An agent, on this view, is any entity with a repertoire that enables it to respond appropriately to what its conditions afford. In turn, the theory of teleological explanation is a recent invariance account (Walsh 2012). The idea is that goals explain the occurrence of the means by virtue of a distinctive relation of counterfactual dependence that holds between them.

The structure of the chapter is the following. In section 4.2 I motivate the ecological approach to agency by showing how such an approach is implicit in our intuitive pre-theoretic understanding of agents. In particular, I argue that the concept of a goal is indispensable to identify the agency of a system as agency. Then in section 4.3 I argue that agency is an empirically observable phenomenon and make explicit the distinctive theoretical role that the concept of a goal plays in the prediction, explanation and evaluation of the behavior of this phenomenon. In section 4.4 I provide a full blown naturalistic account of goals by applying the G2 strategy introduced in the previous chapter. I argue the goals play two distinctive theoretical roles: One in the individuation of agents and the other in teleological explanation. I conclude that this invariance account of

\(^{22}\) In the following two chapters (Chapter 5 and 6) I apply these accounts of agency and teleology to a problem in the philosophy of microbiology and the philosophy of action respectively.
naturalized teleological explanation avoids the objections against natural teleology introduced in Chapter 1.

4.2. The Phenomenon of Agency

4.2.1. The Squirrel and the Acorns

Consider a typical scene in a park. Standing on a tree, a squirrel observes her surroundings. Suddenly she spots something edible nearby. A few meters away a bunch of acorns lie scattered on the ground. She starts running down the tree, sliding from one branch to the other, rushing to take advantage of this opportunity for eating. But as she is about to reach the ground the sudden presence of a dog awaiting her at the bottom impedes her from continuing her descent. She quickly slows down to change her direction towards a long branch that allows her to jump to the next tree to continue her descent and avoid the threat. Once at the ground level she strides forward towards the acorns. But now the landscape reveals new obstacles: A group of pigeons are obstructing the path while a couple of squirrels in the periphery are also advancing towards the food source. So she moves quickly and grabs as many acorns as she can. After putting a few in her mouth, she runs to the shadow of closest tree that provides a secure cover in the face of a potential attack by other squirrels. She digs a hole to hide the acorns and quickly returns to pick up a few more. As she is grabbing the last ones two other squirrels jump aggressively into the scene. But having enough food for herself, she returns back to the shelter of the tree to eat.

4.2.2. The Mechanical and the Ecological

The above story describes the unfolding of the complex dynamic interaction between an organism and its ecological setting. In the previous chapter we saw that the gross dynamics of a complex system may be described in two different yet complementary ways. One is in a fine-grained way in terms of the local or intrinsic properties of its component parts, such as the position and momentum of the individual molecules in a gas. The other is in a phenomenological coarse-grained way in terms of the global or bulk properties of the system as whole, such as the temperature or entropy of the gas. In cases like temperature, entropy and viscosity the fine-grained description captures the conditions that causally realize the system’s gross dynamics. This corresponds to the first stage of G2: showing how some phenomenon is causally realized.
contrast, the coarse-grained description captures the conditions that constitute its (theoretical) identity by specifying its scientific role, in this case thermodynamics. This corresponds to the second stage of G2: specifying the theoretical role of the phenomenon that individuates it.

In the same way, we can understand an organism’s dynamics in terms of the causal organization or architecture that causally realizes its behavioral capacities. Or alternatively, we can understand this dynamics in terms of the set of behaviors that characterize the way it responds to its ecological setting. I will call these descriptions the mechanical and the ecological conceptions respectively. The idea is that just as a fluid emerges as a fluid only when we consider the gross behavior of the whole collection of particles, the agency of an organism emerges only when we consider its gross behavior as ecologically embedded. Each conception serves a different theoretical role and captures a distinctive set of facts about the squirrel. The mechanical conception captures the structural conditions that realize agency in the causal order as per G1-naturalism. The ecological conception captures the behavioral conditions that constitutes agency as per G2-naturalism.

The squirrel is of course a physical object amidst other physical objects. So its behavior is governed by the relevant mechanisms and ultimately by physical laws. In particular, the squirrel possesses a set of physiological and anatomical mechanisms which, when appropriately triggered by external or internal conditions, cause each one of the squirrel’s movements and reactions. Knowledge of these mechanisms allows us to predict how the squirrel will move and to explain how the squirrel’s behavioral response was produced. The behavior of the squirrel can thus be conceived in terms of the activities and capacities of its parts and the way they are organized to produce a behavioral effect given a certain physical input. This sub-organismal description gives us a mechanical conception of the organism’s dynamics.

The mechanical conception specifies causally necessary and/or sufficient conditions for the story of the squirrel in the park to apply. But these conditions do not figure in the content of the story. The story is not about how a complex transduction-device processes physical stimulus to produce a variety of behavioral outputs. Nor is it about how the internal organization of the squirrel is maintained by self-regulatory processes in which matter and energy are exchanged with the environment. The story is rather about the way a purposive agent responds to the network of opportunities and threats that its ecological setting provides relative to its capacities
and goals. That is, about how the squirrel is embedded in its ecological setting. For example, it avoids danger, finds an alternative route to the food, then hides the food and finds shelter. The story thus describes the way the organism as a whole actively responds to the significance of its conditions for attaining its goals. On this description the squirrel is not the passive locus of internal and external conditions. It is an engaged agent directing its behavior in a way that is appropriate to cope with its world. The agential description gives us the ecological conception of the organism’s behavioral dynamics. The squirrel’s capacity to purposively enact the conditions in which it is embedded imbues them with ecological significance. On account of this we cannot describe what its goals are without citing the way its setting matters or is relevant for their attainment. Similarly, we cannot describe the ecological significance of the squirrel’s setting without citing its goals. Thus, the agency of the organism and the relevance of its ecological niche reciprocally constituted each other and together form a single, coupled dynamical system.

As in the case of other complex systems reviewed in the previous chapter, conceiving the squirrel as an agent requires a distinctive set of concepts that are wholly different from the set of concepts used to describe the mechanical causes of its behavior as a physical entity. Specifically, the agency of the squirrel emerges when we describe it from the perspective of its goals such as getting nourishment and avoiding predators; the conditions of its ecological setting that promote or impede the attainment of those goals such as the presence of acorns or the presence of the dog and the shelter of a tree; and its ability to sense and respond to them as promoting or impeding the attainment of its goals, such as escaping, hiding, charging, running, jumping or, digging. These concepts specify not the antecedent conditions that produce the behavior but the conditions for her behavior to succeed and also to some extent how they might fail. They apply not to the parts of the squirrel or its environment but to the whole organism-environment interactive system considered as single, coupled dynamical system. As per G2-naturalism, the gross behavioral dynamics of the organism provides appropriate or prototypical conditions for applying these concepts. As Taylor (2004, 32) puts it

“We have to see the world impinging on these beings in relevance terms; or alternatively put, we have to see them as agents. [...] an engaged agent, determining the significances of things from its aims, needs, purposes, desires.”
4.3. The Theoretical Indispensability of the Ecological Conception

4.3.1. Empirical Basis

The empirical basis of the ecological conception consists of the simple fact that we can see, just by looking at the squirrel’s gross behavioral dynamics as embedded in its ecological setting, what its goals are; what the relevance of its ecological conditions is for attaining them; and that the squirrel responds to those conditions as conducing to the attainment of its goals. ‘Seeing’ consist in a tacit understanding that what we observe fits into a robust, reliably projectable pattern when we characterize under the guise of action. In particular, we can see that the squirrel is avoiding the dog, not merely changing its course. We can see that the squirrel is looking for food, not merely randomly scanning its surroundings. We can see that the squirrel is hiding the acorns, not just depositing them in a hole. We can see that the squirrel runs in the direction of the tree in order to get shelter, rather than just being propelled in that direction. It’s unlikely that we can specify necessary and sufficient conditions for applying these ecological concepts in terms of concepts that belong to the mechanical conception. But even if it were possible there is no need to do so. Ecological concepts form a tight interdefined package that applies holistically to the patterns of behavior that constitutes the unfolding of the squirrel-park dynamics. This behavior thus provides phenomenological evidence for applying ecological concepts. No knowledge of the underlying mechanisms that causally realize the behavior is required, at least for this phenomenon and taking into account that we are defining the theoretical role of agency.

We know that eating or avoiding predators are goals rather than meaningless end-states that the squirrel happens to attain because of the distinctive way by which the squirrel behaves towards acorns and dogs respectively. The squirrel moves towards the food and away from predators. This means that the squirrel will persist in moving in the relevant direction across a wide range of conditions by mounting adaptive changes in its trajectory to compensate for perturbations. For example, if an obstacle such as the dog perturbs the trajectory of the squirrel towards the acorns, the squirrel adaptively changes its trajectory by moving away from the dog and towards an alternative route. The goal-directed character of the squirrel’s responses is thus determined as goal-directed by the way it is embedded in the park, not by its internal organization. That is, by
the relevance that the park’s conditions has for the squirrel given its goals and capacities, not by
the way these capacities or conditions are causally realized.

In the same way, we can see that the squirrel’s behavior is elicited by a stimulus that has
*meaning for* the squirrel, rather than just being caused by the physical properties of the stimulus.
Merleau-Ponty (1942: 31) expresses this point in a quaint way thus,

“[In] a mechanical action [...] the dependence is uni-directional; the cause is the
necessary and sufficient condition of the effect [...] On the contrary ... physical
stimuli act upon the organism only by eliciting a global response which will vary
qualitatively when the stimuli vary quantitatively; with respect to the organism they
play the role of occasions rather than of cause; the reaction depends on their vital
significance rather than on the material properties of the stimuli.”

Furthermore, we can see that a particular piece of behavior fits within the overall pattern of
behavior that the squirrel displays in the park. For example, running towards the acorns and
away from the dog are not just separate physical events. When we consider how they fit in the
overall pattern of squirrel-park interactions their interconnections emerge as a means-goals
dynamics, rather than as a mere cause-effect one. The reason is that the overall pattern exhibits
the distinctive modal profile of hypothetical necessity introduced in the first chapter: The squirrel
would do what is necessary to attain its goals given its circumstances. On the basis of this overall
pattern, it becomes apparent why the squirrel first did the one and then the other rather than the
other way around. The behavioral profile of the squirrel displays a precise order of execution
relative to the relevance of the conditions that elicit the particular responses.

In turn, we know on empirical grounds that the squirrel is responding appropriately or
inappropriately to its conditions because we can see that the squirrel is differentially responsive
to those conditions as being for or against some goal, that is, as providing or impeding effective
means to attain her goals. For example, we can see that the squirrel moves towards the acorns
rather than towards the rocks lying beside them because the former are food for the squirrel and
getting food is her goal. Likewise, we know that the squirrel moves away from the dog rather
than towards it because the dog constitutes danger for the squirrel.
The goals of the squirrel; the conditions that facilitate or impede its attainment as well as the capacity to respond to these conditions as facilitating the attainment of its goals are not discrete entities that we can identify separately from the behavioral dynamics displayed in the relevant situation. They are in the observable pattern of behavior. But they are not localizable independently of the pattern. They are immanent in the robust adaptive dynamics that unfolds between the squirrel and the relevant conditions. In this sense having the goal of collecting acorns, or avoiding dogs is not an internal fact about the squirrel; it is an observable fact about the squirrel as embedded in the park; an ecological fact. The goal of the squirrel emerges as a goal only by considering the whole squirrel-park system as a single, coupled dynamical system. We can’t see a distinctive entity that is the meaning that the acorns has for the squirrel. We can see the meaning that the acorns have for the squirrel in the way the organism behaves towards the acorns. Meaning and purpose are thus constituted as such by the hypothetically invariant dynamics between the organisms and its ecological setting.

One might object that seeing the squirrel depositing acorns in the hole as hiding them is a projection of our own psychological agency into the natural world: there is only the mechanical process of depositing being metaphorically described as the action of hiding. So the ecological conception lacks the mind-independent objective character of the mechanistic conception. One response is to point out that the mechanical conception is as subject to the projectivist interpretation as the ecological conception is. For example, describing biological phenomena such as talk of “molecular machines” in molecular biology is as metaphoric as describing organisms as agents. Seeing or detecting the process of genetic transcription as information processing, that is, as a process of decoding information is a projection of our own engineering capacities into the natural world: there is only the physical process of molecules changing positions metaphorically described as decoding information.

But this objection fails against the ecological conception for the same reason it fails against the mechanical conception. Information talk in molecular biology is justified on the grounds that it is theoretically indispensable. The idea is that a strictly physicochemical non-informational description of the transcription process will fail to capture a distinctive set of empirically detectable and counterfactual-supporting regularities about how this process robustly produces the same functional outcome irrespective of the details of its molecular implementation. In the same way, the view that the squirrel really enacts its ecological setting in a way that subserves its
goals is justified on the grounds that a mere mechanical description of its behavior will fail to capture a distinctive set of empirically observable counterfactual supporting regularities. The transcription process really is an information-processing mechanism at the functional level of description and the squirrel really is a purposive agent at the level of ecological description, just as collection of molecules really is a gas at a thermodynamic level of description. Hence, the fact that the squirrel is hiding the acorns rather than merely depositing them is as much an objective fact as the fact that a gas has a certain temperature. The first is an ecological fact, that is, a fact about the whole organism-environment system. The second is a thermodynamic fact. But neither is a fact about us.

4.3.2. Predictive Role

On the basis of this empirical knowledge we can reliably predict what the squirrel would do next or what the squirrel would have done had her conditions or goals been slightly different. The ecological conception thus allows us to go beyond what we can actually observe to reliably predict what we would observe under a significant range of counterfactual circumstances. For example, just by looking at the squirrel’s behavior we can tell what its goals are but also what its goals would be were the squirrel to behave differently. Likewise, we can tell, just by looking at the squirrel’s surroundings, what features would facilitate or impede achieving its actual or possible goals, and hence which responses would have been appropriate under such conditions. Knowing that the squirrel has the goal of getting nourishment and that there are acorns nearby, we can reliably predict that the squirrel will direct herself towards the acorns. This knowledge allows us to say what the squirrel would have done had circumstances been different. We can predict, for example, that if the acorns are a few meters to the left of where they actually are, the squirrel would direct her behavior accordingly but not if the acorns are, say, a few miles away. Thus appropriateness is strongly predictive. Knowing that the squirrel would respond appropriately, that is, that it would bring about the necessary means to attain its goals given the circumstances, allows us to predict what it would do. If these are genuine predictions then the ecological conception captures them. But there is no guarantee that a mere mechanistic description of the park scene would do so. As the physicist George Ellis (2005, x) puts it
“The human mind [agency] is physically based, but there is no hope whatever of predicting the behavior it controls from the underlying physical laws. [...] Physics by itself cannot explain behavior that is adaptive and depends on context.”

4.3.3. Explanatory Role

A constitutive feature of agents as agents is that they support a distinctive form of explanation, teleological explanation. Teleology is the explanation of the occurrence of some event in terms of the goal, purpose or end that it subserves. To say that the squirrel jumped to the near branch \textit{in order to} avoid the dog is to say that the squirrel jumped to the near branch \textit{because} the squirrel has the goal of avoiding the dog and jumping to the near branch is the appropriate means to attain her goal given her circumstances. Attributing the goal of avoiding the dog explains \textit{why} or for what purpose the squirrel responded in the way she did rather than some other way or no way at all. It does so by showing that the means was \textit{hypothetically necessary} for attaining the goal under the relevant circumstances. This implies that had the goal of the squirrel or its circumstances been different, the squirrel would produce the necessary means to attain it given the new goal or circumstances. So if agency (action) can only be explained in ecological terms and ecological explanations are irreducibly teleological, agency can only be explained teleologically. This explanation contrasts with the explanation of \textit{how} or by which mechanisms the relevant response was causally produced. This indicates that, first, agency plays an indispensable theoretical role and second, that this role is fundamentally teleological. On account of this agency is natural as per G2.

4.3.4. Normative Role

Let us now consider for a moment the distinctive form of explanation that applies to agents as agents, teleology. Teleological explanation has characteristic normative implications. As we saw above goals, \textit{require} their means such that, having a goal imposes a requirement on the agent to bring about the means to its attainment. For example, if the squirrel has the goal of avoiding the dog and the necessary means to avoid the dog is to jump to the nearest branch then the squirrel is required to jump to the nearest branch. Furthermore, on the basis of the goal and the relevant circumstances we can further evaluate the appropriateness or effectiveness of the means for attaining the goal. So if jumping to the nearest branch successfully allows the squirrel to avoid
the dog then we can judge its behavior as appropriate for or good for avoiding the dog. Thus, teleological explanations imply normative requirements on agents and support normative evaluations about the appropriateness or inappropriateness of the means given the agent’s goals and the circumstances. This normativity expresses the fact that agents are characterized by their capacity to respond appropriately to their circumstances guided by norms or requirements that their goals impose. Indeed, it is a matter of empirically observable fact whether hiding the acorns from other squirrels is the appropriate thing to do for the squirrel given its circumstances and goals. Since the normative implications of teleology and hence agency follow from the distinctive set of concepts necessary to capture the distinctive hypothetically invariant modal profile of teleology, these normative implications are unavailable from the mechanical conception. After all, causal-mechanical facts are typically taken to be non-normative. On account of this the ecological conception is indispensable, for it captures a constitutive fact about the nature of some natural entities: the fact that things can be good or bad for them.

We have seen the theoretical indispensability of the ecological conception to predict, explain and evaluate a distinctive set of facts and regularities about organisms. According to a well-established mode of ontological inference (Quine 1969), we ought to be committed to the existence of entities that are theoretically indispensable. It follows that we ought to be committed to the existence of agents. Since agents are constituted by a set of ecological facts, namely, their having goals, encountering opportunities and threats and responding to them as such, it follows that we ought to be committed such entities as goals, opportunities and responding to them as such. But what are goals? What is their place in the causal order of the world? How can they explain anything?

4.4. G2 Naturalization of Goals

According to the G2 strategy of naturalization introduced in Chapter 3, naturalizing a phenomenon x consist of two stages: (i) showing how x is causally realized in the world; and (ii) showing that x plays an indispensable role as x in a scientific theory irrespective of how that role is causally realized. Assuming T1, the principle according to which x is individuated by its theoretical role, it follows that on G2 it is the theoretical role of x rather than the causal realizer of that role that individuates x. In what follows I apply G2 to goals.
4.4.1. The Causal Realization of Goals

With respect to (i) I follow the systems-theoretic account introduced in the Chapter 1 (Sommerhoff, 1950; Bertalanffy, 1969; Nagel, 1977) in first defining the concept of a goal-directed system in behavioral terms and then on that basis derive the concept of a goal. But as per G2, I abandon the reductive aspirations of systems-theoretic approach. That is, I don’t take this account to provide a set of conditions for individuating goals as per stage (ii) of G2, but as providing a set of naturalistic conditions for locating goals in the causal order of the world as per stage (i). The idea is that a system $S$ has a goal $G$ if $S$ can attain and maintain a stable end-state $G$ in a robust (persistent) adaptive (plastic) way, that is, across a wide range of conditions by producing changes in its trajectory to compensate for perturbations. Having a goal is thus an observable gross behavioral, system-level, emergent property of the system (Walsh 2012). A goal, in turn, is simply the stable end-state that a goal-directed system tends to reliably attain (or maintain).\(^{23}\)

As we have seen, goals play two distinctive roles in the case of the squirrel: They allow us to identify it as an agent and to explain its behavior teleologically. That is, on the one hand goals partly individuate agents as agents, and on the other hand we can explain why agents act in the way they do by citing their goals.\(^{24}\) So goals play two theoretical roles. I take each in turn.

4.4.2. The Role of Goals in the Individuation of Agents

We saw how the concept of a goal figures in the individuation of the squirrel as an agent according to the ecological conception of the squirrel dynamics. Before giving a full-blown theoretical specification of what this role is let us consider the theoretical context of the ecological conception of agency.

4.4.2.1. Situated Adaptationism

\(^{23}\) This gross behavioral capacity is in turn a consequence of the architecture of complex adaptive systems, with its characteristic modular and hierarchical network organization (see Kaufman and work in autonomous-systems theory introduced in Chapter 1 (e.g. Barandiaran et al. 2009; Moreno & Mossio 2015)).

\(^{24}\) It is uncontroversial that goals are constitutively necessary for agency. However, I will remain agnostic as to whether there are systems that are not agents that can be said to have goals and hence are susceptible to teleological explanation, such as organismal sub-system such as the immune system or the thermoregulatory system.
The ecological theory of agency forms part of broader attempt to understand the nature of biological organisms as natural agents and its role in adaptive evolution called Situated Adaptationism (Walsh 2013a; 2014b; 2015). Situated adaptationism aims to develop an alternative conception of adaptation in light of the empirical and conceptual objections against the externalist commitments of the standard adaptationist program in evolutionary biology (Lewontin & Gould 1979; Lewontin & Lewins 1987; Lewontin 2000). This alternative conception is also motivated by recent empirical and theoretical developments in evolutionary developmental biology and organismal biology more generally that are based on the application of complex systems dynamics principle and models to biology (Kauffman 1993; Kirschner & Gerhart 2005; Noble 2006; West-Eberhard 2003). What this research indicates is that rather than being mere “survival machines” passively controlled by their genes and the environment as per the standard genocentric approach (Dawkins 1983), organisms actively engage their conditions of existence to respond appropriately to the internal and external challenges they face. Organisms are, in short, agents rather than patients of their own adaptive responsiveness (Kauffman 2000; Walsh 2013a; 2015). In order to vindicate the constitutive role that organisms play as adaptive agents in determining their conditions of existence and hence in responding adaptively to these conditions, following Gibson (1979) the situated adaptationist approach proposes to conceive the environment or ecological niche as a set of affordances or possibilities for engagement. To conceive the environment as an “affordance landscape”, in turn, implies that organisms are goal-directed entities capable of responding appropriately to the relevance of their conditions of existence for fulfilling their life cycle. Thus, the purposive character of organisms and the relevance of their ecological niche reciprocally constituted each other and together form a single, coupled dynamical system. In the words of Walsh (2013a, 267):

“For a system to experience its conditions of existence as affordances it must generally be capable of responding to them as affordances, by exploiting the opportunities they provide for the attainment of its goals or by mitigating the impediments. In turn, to be a purposive system is to be a system that is capable of responding to its conditions in a way that is conducive to the fulfillment of, or the amelioration of the impediments to, those goals.”

In what follows I will develop some of the central insights of situated adaptationism about the nature of agency. The central idea is that to characterize agency as an ecological phenomenon we need to appeal to three inter-definable concepts: goal, repertoire and affordance. Together, these
concepts specify the ecological profile of the system. To be an agent is to exhibit an ecological profile. The concept of an agent is thus irreducibly ecological. An agent’s goals are constrained by its affordances and repertoire. The idea is that these ecological concepts are necessary to specify what the goals of an agent are. I take each one in turn.

4.4.2.2. Goals

A system $S$ has a goal $G$ if $S$ has the behavioral repertoire to attain and maintain $G$ across a wide range of actual and counterfactual conditions $C$ by responding appropriately to what $C$ affords. If $G$ is $S$’s goal and $r$ is a means to the attainment of $G$ in conditions $C$, then ceteris paribus, $S$ would bring about $r$ in $C$ for a suitably large range of values for $G$, $r$ and $C$. And if $r$ is detrimental to the attainment of $G$ in $C$, then $S$ typically won’t bring about $r$. A system $S$ has a goal $G$ then if it supports the following counterfactuals: If the goal $G$ were to change to $G^*$ while conditions $C$ remain constant, then $S$ would modify its behavior $r$ accordingly towards $G^*$ by producing $r^*$. And if the conditions $C$ change to $C^*$ while the goal $G$ remains constant, then $S$ would modify its behavior $r$ accordingly to maintain its trajectory towards $G$ by producing $r^*$. I call this modal profile hypothetical invariance. This profile underlies the predictive, explanatory and normative implications of the ecological conception. A state of affairs $G$ constitutes a goal then if it can be attained and maintained in a hypothetically invariant way. Since hypothetical invariance is a complex modal relation that holds between $S$ and $C$, a state of affairs $G$ emerges as a goal only when we consider the whole $S$-$C$ system as a single coupled dynamical system.

For the squirrel to have the goal of getting nourishment is for the squirrel to behave in a hypothetically invariant way towards whatever condition in the park affords nourishment, such as the acorns. This implies that if the trajectory of the squirrel towards the acorns were blocked by some pigeons, the squirrel would modify its behavior appropriately, for example, by changing its direction or by doing something that would make the pigeons disperse. And if the goal of the squirrel changes from getting nourishment to avoiding a predator, then the behavior of the squirrel would change accordingly in an appropriate way from running in one direction to running in the opposite one. Eating the acorns constitutes a goal for the squirrel by virtue of the gross behavioral capacity of the squirrel to respond to the acorns as edible, that is, by virtue of its capacity to attain and maintained its trajectory towards the acorns in a hypothetically invariant way. The property of having a goal $G$ is thus neither localized or instantiated inside the agent as
a representational structure nor outside it in $C$ as a self-standing state of affairs with intrinsic value. Rather it is immanent to the behavior, that is, it is in the behavioral pattern that emerges from the interaction between $S$ and $C$ and hence it is dynamically constituted by such interaction.

4.4.2.3. Repertoire

Hypothetical invariant behavior implies a repertoire, that is, a set of responses that the system can produce to current conditions and the capacity to select the sub-set that is appropriate given what its conditions afford. A repertoire is the set of things that the system can do or the set of purposive behaviors it can generate. A repertoire is not mere set of movements but rather is constituted as a repertoire only with respect to some goal $G$ and some set of conditions $C$ that facilitate (or impede) the attainment of $G$. Given its purposive character, a given response from a repertoire must be selected by the system because it constitutes an appropriate means to attain (or maintain) $G$ in $C$.

The capacity of $S$ to bias $R$ towards a response that is conducive to the attainment of a goal is a system-level emergent property of the system as a whole. It consists of a relation between the system as a whole and the activities of its constituent parts. These parts themselves possess a repertoire of possible activities that the system as a whole regulates so as to induce those activities that are conducive to the attainment of a goal (Walsh 2013a). This self-regulatory capacity of the system as a whole is underwritten by the distinctive modular-hierarchically arranged architecture of complex adaptive self-organizing systems (Kauffman 1993; Simon 1962). The capacity to select an item from the repertoire thus need not amount to a decision in the rational sense of being the outcome of deliberation. Rather, it just the gross capacity of a goal-directed system to bias its capacities in a goal-conducive way, that is, in way that exploits or mitigates what its conditions afford for attaining it goals.

Consider the repertoire of the squirrel. It includes items such as jumping, climbing, running, eating, hiding, escaping and grabbing. What makes this items part of the repertoire of the squirrel is the fact that the squirrel not just randomly produces any of them on a given occasion. Rather, the squirrel tends to produce the appropriate item in response to the relevant opportunities or obstacles (affordances) it encounters. As we saw in sub-section 4.3 the squirrel doesn’t just move in an opposite direction from a given environmental stimulus, say the presence of the dog. The
squirrel runs away or *escapes* from *danger*. ‘Escaping’ and ‘danger’ are ecological concepts. Given the appropriate condition, each of these responses constitutes a means to attain a goal. The squirrel’s repertoire is reciprocally constituted as a repertoire by the fact that the park offers a set of opportunities and threats that in turn is reciprocally constituted by the fact that the squirrel has a set of goals. So jumping, climbing, and the like are no mere mechanical movements but purposive responses to affordances and hence ecologically significant features of the organism’s surroundings. The concept of a repertoire is meant to capture this ecological fact. To have a repertoire is to have the gross capacity to respond to conditions in a certain (adaptive) way, namely, as opportunities or threats. But what is for a set of conditions to offer or promote or facilitate the attainment of a goal $G$?

### 4.4.2.4. Affordances

A repertoire implies a set of conditions that facilitate or impede the organism to attain its goals. These conditions are affordances. Affordances are opportunities/threats or possibilities/impossibilities for action that the causal structure of an organism’s external or internal conditions of existence provide (or impede) relative to its repertoire and goals.

According to Gibson (1979, 127): “The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill”. Affordances are indexed to organisms. Acorns afford food to squirrels but not to whales; oxygen is poison for anaerobes but not for aerobes, so oxygen is a negative affordance to anaerobes and positive affordance for aerobes. A hole in the wall affords escaping to a mouse but not to a horse; and water affords drinking to most animals, swimming to a fish and a supportive surface to water striders. A rock affords different things to a spider and to a primate. To the spider it is a shelter, to the primate it is a tool. In turn, organisms are indexed to affordances. Being under water affords a fish the possibility of performing vital activities such as movement, respiration, nourishment, etc. while being outside water impedes their appropriate performance. As this reciprocal indexicality indicates, the role of the concept of an affordance is to capture the constitutive interdependence or “complementarity” (Gibson, 1979) or “commingling” (Haugeland, 1998) between organism and environment, the fact that there cannot be one without the other. In Gibson’s (1979, 8) words,
“[T]he words animal and environment make an inseparable pair. Each term implies the other. No animal could exist without an environment surrounding it. Equally, although not so obvious, an environment implies an animal (or at least an organism) to be surrounded.”

What a set of conditions affords depends on what the organism can do, its goal-biased repertoire. And what an organism can do in turn depends on what its setting affords. For example, if a surface is rigid relative to the weight of an organism with the goal of laying down it affords support; if the length of a stair is proportional to the length of the leg of a bipedal organism with the goal of going up it affords climbing; and if the distance between two trees is close enough relative to the size, weight and propulsion capacities of an organism with the goal of getting from one tree to the other it affords jumping. Since affordances are partly constituted as affordances by the goal-bias repertoire of an organism, they serve as a prompt for responding appropriately or inappropriately. So ceteris paribus, the organism will respond to surfaces as supportive, to distances as jumpable and to stairs as climbable. And if squirrel in the park has the goal-biased repertoire of getting nourishment, there are a bunch of acorns scattered in the ground and acorns constitute nourishment for the squirrel, then the acorns afford nourishment to the squirrel. Hence, ceteris paribus, the squirrel will respond to acorns as edible by deploying from its repertoire those items that are conducive to their attainment.

In short, an organism considered as a purposive agent and its conditions of existence considered as a network of affordances reciprocally constitute each other a together constitute a single, coupled dynamical system (Walsh 2014b). This implies that a change in the affordances implies a change in the goal-directed capacities of the organism and vice versa. Following Walsh (2011) we can call the complete set of affordances that an organism encounters its affordance landscape. To be an agent is to inhabit and co-construct an affordance landscape.

There is considerable controversy about the metaphysics of affordances. Two main views are dispositionalism and emergentism. Roughly, the former holds that affordances are dispositional properties of the environment that are triggered by the reciprocal dispositional properties of an animal to respond to them (Scarantino 2002; Turvey 1992). According to the latter, affordances are emergent entities jointly constituted by the capacities of the organisms and its physical surroundings (Chemero 2003; Stoffregen 2003; Walsh 2011). Instead of being the sum of
environmental and organismal properties as dispositionalism claims, emergentism conceives affordances as qualitatively different from the organismic and environmental properties that realize them. Affordances are thus irreducible functional relations between the organism as a whole and the relevant features of the environment.

The problem with dispositionalism is that the environment instantiates the relevant dispositions regardless of whether there is an organism present to trigger them or not. Thus, dispositionalism fails to accommodate the essential complementarity between organism and environment. Rather, what make affordances explanatorily salient are the triggered or ‘actuated’ possibilities. Affordances are not dispositions, which are intrinsic modal properties, but mutual or reciprocal manifestations of possibilities. But realized possibilities are actual relations between the organism and the relevant condition. Thompson (2007, 74) captures this relational-emergent nature when he says,

“While sucrose is a real and present condition of the physicochemical environment, the status of sucrose as a nutrient is not. Being a nutrient is not intrinsic to the physicochemical structure of the sucrose molecules; it is a relational feature, linked to the bacterium’s metabolism. [...] Sucrose has meaning and value as food but only in the milieu that the system itself brings into existence or constitutes for itself.”

In the same way, being nutritious is not an intrinsic property of the chemical structure of the acorns. Although it is by virtue of its chemical structure that the acorns have the potentiality to constitute food for an organism with the right digestive capacities and goals. Being nutritious is rather an ecological relation between the acorns and the squirrel. Or more precisely, it is a relation between the manifested dispositions of the acorns together with the manifested disposition of the squirrel to eat them. The property of being nutritious is thus an emergent relational property jointly constituted by the chemical structure of the acorns and the goal-biased capacities of the squirrel to eat. It is a property of the whole squirrel-acorn system qualitatively distinct from their constituent parts. An affordance is thus causally realized by the respective contributions of the relevant dispositions. But it is individuated as an affordance in terms of the manifest purposive responses they facilitate or impede in attaining a goal.

As we saw in the first section of this chapter, the ecological conception is committed to the view that organisms respond to relevance and hence to the significance or value that things have for
attaining their goals. Relevance is thus a normative notion and the concept of responding to an affordance, offers a naturalization of this notion. As Gibson said, affordances are “for good or ill”. Insofar as affordances are real objective features of the world—organism-environment relations—the normativity they support is also real and objective. In the words of Gibson (1979, 33; 1982, 407),

“The world of physical reality does not consist of meaningful things. The world of ecological reality ... does. If what we perceive were entities of physics and mathematics, meanings would have to be imposed on them. But if what we perceive are the entities of environmental science, their meanings can be discovered.”

“The meaning and value of a thing consist of what it affords”. 25

25 See Sanders (1993)

The normative force of responding to affordances can be articulated in the following terms: if a condition $x$ of an ecological setting $C$ affords $G$ to a system $S$ by specifying the necessary means $ri$ for attaining $G$ given $S$’s repertoire $R$, then two normative implications follow: (1) $G$ requires $S$ to do $ri$ and (2), doing $ri$ can be evaluated as an appropriate or inappropriate or good in light of $G$ and $x$. Consider the first implication. Typically, if a system $S$ has goal $G$ in $C$ then ceteris paribus $S$ would tend to behave in a hypothetically invariant way towards $G$ by responding appropriately to what $C$ affords. If condition $x$ of $C$ affords $ri$ and $ri$ is the necessary means to attain $G$ in $C$ then, ceteris paribus, the fact that $x$ affords $ri$ and $ri$ is conducive to $G$ imposes a requirement on $S$ to bring $ri$ about. Responding to affordances exhibits the distinctive model profile of hypothetical invariance. This modal profile constitutes the normative profile of responding to affordances. The fact that $S$ responds to $x$ as affording $ri$ is then an instance of norm-guided behavior. The ecological approach allows us to vindicate the normative character of purposive agency as normative without jeopardizing its naturalistic status and hence offers a non-reductive (emergentist) form of natural normativity.

Now consider the set of (axiological) implications of affordances. If $x$ is an affordance then there is such a thing as an appropriate response to $x$. This means that if $x$ is an affordance for $G$, then $x$ is good (or bad) for or significant for $G$ and hence $x$ has value for $S$. To experience one’s
ecological setting as an affordances landscape is for the elements of that setting to have value of significance, that is, to be relevant for the attainment of one’s goals. So the ecological account allows us to say that values are real objective properties of the world—ecological relations—and wholly natural as values. Goodness, significance and value are not intrinsic properties but emergent relational properties reciprocally constituted by the goal-biased repertoire of $S$ and what its setting $C$ affords. These normative properties are thus both, causally realized and distinctive normative as per G2-naturalism.

Normative facts such as the fact that agent $S$ is required to do $r$ in response to $x$ or that doing $r$ is good for $G$ or that doing $r$ has value for $S$ are not facts about $S$ per se or of $r$ per se or of $x$ per se taken in isolation. Rather they are facts about the way $S$ is embedded in its ecological setting $C$, that is, in its affordance landscape. This implies that normative facts are constituted as normative by the whole $S$-$C$ system. Thus, although it may be true of agent $S$ that she ought to do $r$, and it may be true of action $r$ that it is the rational thing to do, and it may be true of $x$ that it affords $r$, these normative claims are made true by the dynamics of the highly complex overall ecological system $S$-$C$.

Let us return to the example of the squirrel. The physical description of the relation between acorns and the squirrel explains how the purposive activity of getting nourishment is produced and realized in the causal order of the world. But it does not capture the distinctive hypothetically invariant relation that constitutes the squirrel as an agent and hence explains why it behaves in the way it does and how it would behave and how it should behave. The ecological conception as theoretically developed in terms of the concept of a goal, repertoire and affordance, thus play an indispensable role in capturing a distinctive set of empirically observable and counterfactual supporting regularities about the overall dynamics of the whole organism-environment system. So even if the conditions for hypothetical invariant behavior can be specified in physical mechanical terms, their distinctive explanatory, predictive and normative content cannot. They must be specified in terms of the goals and repertoire of an agent as embedded in its affordance landscape. On account of this agency is natural as per G2.
In summary, agency is a gross dynamic property of a purposive system embedded in its ecological setting. It can be explicated in terms of three inter-defined concepts—goals, repertoire and affordance.

**Goal def** = A state of affairs $G$ is a goal only if there is a system $S$ that tends to attain and maintain $G$ by marshaling its repertoire $R$ in response to its affordances.

**Repertoire def** = A repertoire $R$ is a biased range of potential behaviors $r \{r_1, r_2..., r_n\}$ that enables a system $S$ to realize its goals $G$ in response to its affordances.

**Affordance def** = An affordance $x$ is a property of an organism-environment system that impedes or promotes the deployment of a system $S$’s repertoire $R$ in pursuit of its goals $G$.

### 4.4.3. The Role of Goals in Teleological Explanation

We have seen the role that goals play a theoretical ecological account of what an agent is. This theoretical role is, in accordance with G2, specified in terms of the ecological concepts of ‘affordance’ and ‘repertoire’, not in terms of the concepts that specify the conditions that realize goals in the causal order of the world. Now let us consider the role that goals play in teleological explanation. As we will see, the specification of the explanatory role of goals in teleology appeals to the concepts of invariance and conduciveness.

In Chapter 1 we saw some problems with the idea of goals explaining as goals. Here is a solution: invariance. According to a prominent account of scientific explanation in the philosophy of science, explanation is an inherently modal notion. An explanation provides an answer to questions about what-if-things-had-been-different (Woodward 2003, 211). The idea is that to explain the occurrence of $x$ in terms of $y$ is to cite an invariance relation between $x$ and $y$. An invariance relation is a robust counterfactual relation of dependence between $x$ and $y$ such that if the value of $x$ were to change while maintaining background conditions constant then the value of $y$ would change accordingly in a systematic way (but not vice versa). The invariance

---

26 See also Reed (1996) and Withagen et al. (2012).
account of scientific explanation was originally introduced to account for causal explanation. But it has been argued recently in the philosophy of biology that there is a distinctive invariance relation between a goal (effect) $y$ and means $x$ (cause) that renders teleology as respectable as causal explanation (Walsh 2008; 2013b). The basic idea is that in a teleological explanation the relation of invariance is the inverse from a causal-mechanical explanation. Means counterfactually depend on their goals. They do so in a way that is analogous (structurally equivalent) to the way that effects counterfactually depend on causes. If invariance is the relation that enables causes to explain their effects, and the same relations holds between goals and their means, then we should be able to explain means by appeal to goals in the same way that we explained effects by appeal to their causes. So by parity of reasoning, if causal explanation is legitimate form of scientific explanations then teleology is too.

Walsh proposes to conceive explanation along two dimensions, a cognitive and a metaphysical dimension. The reason for the cognitive dimension is that explanation is meant to provide knowledge or understanding about the world. So the description of the relations between explanans and explanandum must be such that it is made intelligible. And the reason for the metaphysical dimension is that explanation is meant to be about the world. In the case of causal-mechanical explanation, the metaphysical condition is satisfied by identifying a relation in which the effect $y$ counterfactually depends on the cause $x$ such that changing the value $y$ while maintaining background conditions $C$ constant changes the value of $x$ in a systematic way. The characteristic counterfactual implication is that if $x$ hadn’t occurred $y$ would not have occurred either. The cognitive condition, in turn, is satisfied by describing the relation between $x$ and $y$ as a producing relation. The description makes clear what kind of relation we are dealing with. It is not just counterfactual dependence. It is pulling, twisting, attracting, repelling, etc. It tell us what is about $x$’s parts and activities that brings about $y$. This requires a distinctive set of causal concepts including dispositional concepts.

In turn, in the case of teleology the metaphysical condition is satisfied by identifying a relation in which the means (cause) $x$ counterfactually depends on the goal (effect) $y$ such that changing the value $y$ while maintaining background conditions $C$ constant changes the value of $x$ in a systematic way in the sense that if the system is question had had different goals while the background conditions remain constant, it would have produced a different behavior. Call this hypothetical invariance. Hypothetical invariance is characterized by the set of counterfactuals
that it supports: If goal \( y \) were to change to \( y^* \) while conditions \( C \) remains constant then some other means \( x^* \) would have occurred that would had caused \( y^* \); and if goal \( y \) were held constant while \( C \) changes to \( C^* \) then some other means \( z \) would have occurred that would had caused \( y \).

The cognitive condition, in turn, is satisfied by describing the relation between \( x \) and \( y \) as one where \( x \) is conducive to \( y \), where \( x \) is conducive to \( y \) just in case \( x \) is a causally effective in bringing \( y \) about given \( C \), and \( y \) requires \( x \) in the sense that \( x \) is necessary for attaining \( y \) given \( C \). In turn, if means reliably causes a goal it is good for attaining that goal. Thus, the notion of conducing has a hybrid Janus face character: simultaneously causal and normative. So the invariance account of teleology vindicates the distinctive modal and normative profile of the goals-means relation. The justification is that we have to describe the invariance relation in a way that discloses the fact that the effect is a goal and the means is a reliable way of achieving it given the circumstances. We use locutions like ‘for the purpose’, ‘for the sake of’ and ‘so that’ to signify that this is a conducive relation. This implies that the means causes, if successful, the goal. But the goal explains the occurrence of the means. Thus, we cannot replace the ‘conducing’ relation with the ‘producing’ relation without eliminating crucial information. So by virtue of being an instance of two different regularities elucidated by two different kinds of descriptions, the event of hiding acorns can be explained twice over. Thus just as the mechanical conception of an organisms dynamics, the ecological conception supports a distinctive form of explanation.

We can now see that the problem with the eliminativist accounts of teleology reviewed in the first chapter. They took for granted that if teleology constitutes a genuine form of explanation at all then it must be an instance of some kind of causal-mechanical explanation in disguise. But by taking invariance rather than causation as the primitive notion in a theory of explanation, the invariance account of teleology avoids eliminativism. In turn, the problem with the primitivist account of teleology reviewed in the first chapter is their commitment to the view that, since teleology constitutes a genuine non-mechanical mode of explanation by virtue of its distinctive content, then causal-mechanical relations don’t figure in teleological explanations. But the relation between means and goals is a causal relation and hence presupposes that the relevant mechanisms are operative. However, since it is goals that explain the means by hypothetically

\[ \text{\footnotesize 27} \] So far I have used the notion of ‘conduciveness’ without defining it. For example, the repertoire of an agent is such that it can be biased to those activities that are conducive to the attainment of the goal.
necessitating them, the causal-mechanical relation doesn’t figure in the content of the explanation. So while eliminativist is right that teleology applies under the right causal-mechanical conditions, it is wrong in supposing that therefore these conditions figure in the content of the explanation. And while primitivism is right that using strict causal-mechanical concepts cannot specify the content of teleological explanations, it is wrong that therefore no causal-mechanical conditions can be specified for applying teleology. So the invariance account of teleology provides a framework for preserving the good while avoiding the bad of the standard accounts.

4.5. Conclusion

The invariance account of teleological explanation effectively responds to the objections against natural teleology introduced in the first chapter: The alleged commitment of teleology to causation by non-actualia, intentionality, normativity and the view that the completeness of mechanism renders teleology dispensable.

The objection from causation by non-actualia holds that since goals are future unactualized state of affairs teleological explanation is committed to a mysterious form of causation incompatible with the physical efficient causation familiar from natural science. But the invariance account of explanation denies the assumption that that for \( x \) to explain \( y \) is for \( x \) to cause \( y \). The central explanatory relation is invariance. Causation is a particular instance of an invariance relation but not the only one. Goals explain their means not by causing them. After all, it is means that reliably cause their goals. Rather, goals explain their means by requiring them, in the sense that means are hypothetically necessitated by their goals and circumstances, where the modal and normative profile of hypothetical necessity is captured by the concepts of ‘(hypothetical) invariance’ and ‘conduciveness’.

The objection from intentionality claims that the only way to avoid causation by non-actualia is to take intentions, the mental representation of goals, rather than goals themselves as explananda. But since the invariance account of teleology prevents the objection from non-actual causation that motivates the objection from intentionality, the motivation for this objection from intentionality dissolves. Moreover, the notion of intentionality does not figure in the behaviorist account of how goals are causally realized in the world or in the specification of the theoretical
role they play in the individuation of agents and in teleological explanation. As Aristotle (1970, 8) put it, “it is absurd to suppose that purpose is not present because we do not observe the agent deliberating”. More recently Dreyfus (1992, xxxi) has argued that “behavior can be purposive without the agent having in mind a goal or purpose”. I have argued that this is an empirically observable fact about the way certain entities such as organisms are embedded, like the example of the squirrel in the park indicates.

The objection from normativity claims that for goals to require their means and hence for them to be such that they ought to be attained goals must be intrinsically evaluable state of affairs (Bedau 1992). But according to the invariance account of teleology an agent is normatively required to bring about the means to attain its goals whether or not the goals ought to be attained (Broome 1999).\(^{28}\) This relation furthermore is captured in terms of the notion of conduciveness—means cause their ends reliably—which is simultaneously causal and normative (means are good for/appropriate for their ends). So not only there is no need to posit intrinsically evaluable state of affairs, the relevant normative relation is causally realized.

Finally, the objection from explanatory exclusion claims that on account of the completeness of mechanism, there is no role left for goals to play in explanation (Kim 1989). But this objection presupposes, again, that for \(x\) to explain \(y\) is for \(x\) to cause \(y\) such that goals would have to cause in a non-mechanical way for them to be explanatory. But again, the mark of explanation is invariance, not causation. Mechanical explanations instantiate a distinctive invariance relation, namely, causal productivity (the effect counterfactually depends on the cause). But teleological explanations instantiate a different invariance relation, namely, conduciveness (the cause counterfactually depends on the effect). Since mechanical and teleological explanations are structurally isomorphic, and we already accept mechanical explanation, then by parity of reasoning teleological explanation should be accepted as a distinctive form of explanation (see Walsh 2012).

The ecological account of agency and the invariance account of teleology offer a G2 naturalization that avoids the dichotomy between eliminativism and primitivism giving us a

\(^{28}\) Normative requirements are hypothetical but obligations are categorical.
middle way between these two extremes that allows us to satisfy both, naturalism and explanatory (normative) adequacy.
Chapter 5
The Case of Bacterial Cognition

5.1. Introduction

In the previous chapter I introduced the ecological account of natural agency and an invariance account of teleological explanation as an instance of G2 naturalism. The purpose of this and the next chapter is to apply the methodology, diagnosis and positive proposal to areas of philosophy where the naturalization of a normative phenomenon is question. In this chapter I contrast the ecological account of agency with the traditional Cartesian conception using recent research in Bacterial Cognition and Cellular-decision-making as a test case. I argue that the Cartesian conception forces us to choose between attributing full-blown intentional psychology to bacteria and treating them as mere machines. Most organisms, however, occupy a middle ground between these two extremes: On the one hand their behavior is too complex and supplely adaptive to count as mechanistic. On the other hand it appears to fall well short of full-blown cognition. So Cartesianism faces a dilemma between mechanism and intellectualism. An alternative ecological conception of agency as responsiveness to affordances, I argue, allows for degrees of agency along a continuum, from the simplest adaptive agents such as bacteria to the most sophisticated cognitive agents such as human beings. Ecological agency thus resolves the Cartesian dilemma.

Recent empirical work in bacterial cognition and cellular-decision-making has exposed the remarkable capacities of the simplest livings systems to respond appropriately across a wide range of conditions. To make sense of these capacities, the distinctive conceptual strategy of this research has been to use cognitive and other psychological concepts, hence the names “cellular decision making” or “bacterial cognition”. This strategy is not typically meant as a mere metaphor or heuristic device, but as an accurate characterization from which novel theoretical implications can be drawn about the nature of life and mind.

I argue that the cognitive turn in microbiology is predicated on the Cartesian assumption that goal-directed agency is the mark of cognition. This view, in turn, generates a dilemma between mechanism and intellectualism: A system is either a mere automaton or a full-blown cognitive
agent; there is no middle ground. But this conceptual scheme, I argue, is not sufficiently nuanced to capture what these important empirical discoveries mean for our understanding of the nature of agency. While mechanism fails to capture bacteria as agents, intellectualism succeeds but by overestimating their agency as rational. We need a via media between these two extremes. The ecological account of agency, I argue, gives us this via media. According to this view, an agent is an ecologically embedded purposive system. That is to say, any system with the gross behavioral capacity to bias its repertoire in response to what its conditions afford for attaining its goals. Since the repertoire of a system can vary in its richness, I argue, there can be degrees of natural agency along a continuum, from mere adaptive agents at one extreme to cognitive agents on the other. Thus, goal-directed agency does not imply cognition. The ecological account vindicates what we might think of as the common sense solution to the problem: it seems to exert irreconcilable demands on a theory. Our account of agency must show both that bacteria are adaptive agents, hence not mere automata, but not cognitive (rational) agents.

5.2. The Cognitive Turn in Microbiology and Cell Biology

5.2.1. Bacteria are Purposive Agents

Recent empirical research in Bacterial Cognition and Cellular-Decision-Making has exposed the remarkable capacities of the simplest living systems to respond appropriately across a wide range of conditions. For example, cells can attain and maintain functionally important states spontaneously, that is, “without an associated genetic or environmental difference” (Balážsi et al. 2011, 910). Genetically identical populations of bacterial cells sharing the same environment can display markedly different phenotypes (“fates”). This may be the consequence of chance due to internal stochastic fluctuations (Choi et al. 2008), external signals triggered by environmental changes or an “autonomous decision” (Norman et al. 2013). The cell is able to integrate a variety of internal and external stochastic signals from past and present conditions (Losick & Desplan 2008) to anticipate future conditions and respond accordingly (Tagkopoulos et al. 2008; Perkins and Swain 2009). Bacterial cells can also detect and repair lesions to their DNA (Shapiro 2007; Ciccia & Elledge 2010) and sense the density of the population (“quorum sensing”) in order to
form colonies ("biofilms") that enable them to respond collectively as a multicellular organism (Ben-Jacob 2005).  

More famously, unicellular organisms possess a variety of motile capacities or “taxes” which are sensitive to various kinds of directional gradients of stimulus, such as light, gravity, temperature, oxygen and magnetic field (Bray 2000). Consider chemotaxis. This is the capacity of bacteria to direct their movement towards chemical compounds that are nutrients or “attractants”, such as glucose, and away from chemical compounds that are toxins or “repellents”, such as phenol, by sampling the concentration of these chemicals in their environment. Chemotaxis results from alternating two kinds of movements, tumbling at random and swimming in a straight line. If the bacterium senses that it is moving towards the attractant and away from the repellent, it will keep swimming in a straight line for a longer period before it tumbles. But if it senses that it is moving towards the repellent and away from the attractant it will tumble sooner in order to try a new direction at random. Chemotaxis is thus achieved by modulating the tumbling frequency.

The signal-transduction-pathways, biochemical mechanisms and regulatory networks that cause and realize and control chemotaxis are well known (Porter et al. 2011). Roughly, chemotaxis occurs when extracellular chemical signals are detected by transmembrane receptors that transduce the signal to flagellar protein motors. These motors control the cell’s movement by rotating either counterclockwise for swimming in a straight line or clockwise for tumbling at random in response to feedback. This sensorimotor causal organization enables the self-regulatory capacity of the bacterium to attain and maintain a set of homeostatic stable end-states in a robust adaptive way, that is, across internal and external perturbations by making compensatory changes to its trajectory (Alon et al. 1999). Chemotaxis thus involves the coordination of various subsystems by the cell as a whole. By means of this signal-transduction architecture, the cell as a whole is able to produce and sustain a supple adaptive dynamics with its (internal and external) environment.

This robust supple adaptive dynamics is the mark of goal-directed behavior (Rosenblueth et al. 1943; Sommerhoff 1950; Bertalanffy 1969; Boorse 1976; Nagel 1977; Walsh 2011). Goal-directed behavior, we saw in the previous chapter, is characterized by its distinctive modal 

---

29 For a comprehensive review of this research see Lyon (2015).
profile: If some end state $G$ is a goal, then the system would produce the necessary means to attain and maintain $G$ across a wide range of actual and counterfactual circumstances. In this case it is the whole bacterium rather that a sub-system that is the appropriate unit to which to attribute the goal $G$. Because of this robust counterfactual dependence between means and goals, knowing that a system’s goal is $G$ allows us to reliably predict and explain what the system would do under a range of conditions (Walsh 2011). The sensorimotor organization of the cell thus supports a distinctive set of empirically observable and counterfactual-supporting purposive or goal-directed regularities. But because of their distinctive modal profile of these purposive regularities are insensitive to details of the mechanisms that underpin them.

Moving towards the source of attraction is not something that just passively happens to a bacterium as a blind mechanical consequence of genetic and environmental events. Attaining and maintaining the goal state is something that the system actively does by marshaling the resources of its repertoire in response to its affordances. In the case of the bacterium it is something that the bacterium actively does by exploiting (or mitigating) its genetic and environmental circumstances in order to fulfill its life cycle. Hence, it is something that the bacterium can achieve or fail to achieve as a whole on account of which we can evaluate the appropriateness of the bacterial response given the organism’s goals and circumstances. Bacteria are thus purposive agents guided by the demands that the chemical content of its environment imposes for fulfilling its biological goals. This should not be surprising given that the purposive character of living systems is typically a if not the distinctive feature of livings systems as such. As Sommerhoff (1950, 6) puts it,

“On the phenomenal level from which all science must proceed, life is nothing if not just this manifestation of apparent purposiveness and organic order in material systems. In the last analysis, the beast is not distinguishable from its dung save by the end-serving and integrating activities which unite it into an ordered, self-regulating, and single whole, and impart to the individual whole that unique independence from the vicissitudes of the environment and that unique power to hold its own by making internal adjustments, which all living organism posses in some degree.”
We have the causal-mechanical concepts to explain how goal-directed capacities emerge as the effect of the activities of underlying biochemical mechanisms (i.e. feedback loops, circuits, pathways and networks). But what kind of concepts should we use to make sense of these goal-directed agential capacities *as such*, that is, as underwriting teleological explanations and normative evaluations? In the previous chapter it was argued that the relevant concepts are ecological. But this is not the strategy of this research.

### 5.2.2. Purposive Agency is the Mark of Cognition

One way to make sense of these capacities is to think of them in cognitive or psychological terms. This is the strategy taken by researchers working on bacterial cognition and cellular-decision-making. Psychological agents have goals—their intentions or desires—and act according to norms—the norms of reason. Goal-directed agency is hence the mark of cognition, the psychological capacity for rational thought and action. McClintock’s (1984, 184, 193) path-breaking description of genome repair is perhaps the earliest example of this strategy. She begins by characterizing cells as goal-directed systems,

> “The conclusion seems inescapable that cells are able to sense the presence in their nuclei of ruptured ends of chromosomes, and then to activate a mechanism that will bring together and then unite these ends, one with another. And this will occur regardless of the initial distance in a telophase nucleus that separated the ruptured ends. This ability of a cell to sense these broken ends, to direct them toward each other, and then to unit them so that the union of the two DNA strands is correctly oriented, is a particularly revealing example of the sensitivity of cells to all that is going on within them.”

And she ends by characterizing them as cognitive agents:

> “They make wise decisions and act upon them. [...] A goal for the future would be to determine the extent of knowledge the cell has of itself and how they utilized this knowledge in a “thoughtful” manner when challenged.”

Following McClintock’s original insight, there has been a tendency to implicitly infer cognition from goal-directed agency. Baker and Stock (2007, R1023), for example, envisage the next task
of research on bacterial signal-transduction to be that of understanding “how these elements [genes and proteins] are connected to form a dynamic, adaptive cell”. A task they conceive in cognitive terms:

“How is information converted into knowledge, and how is knowledge sorted, evaluated and combined to guide action, morphogenesis and growth?”

Likewise, Shapiro (2007, 809, 816) talks about the “purposeful action by cells under challenge” as motivating the “contemporary view of cells as cognitive entities acting in response to sensory inputs”. On this view “bacteria continually pick up and process information about the environment, internal conditions and other cells to decide on appropriate biochemical and biomechanical actions.” Similarly, Ben-Jacob et al (2006, 504), a leading expert in biofilm formation, conceives “the colony’s capabilities to perform collective sensing, distributed information processing and collective gene-regulation as fundamental cognitive functions.” And Bray (2012), a leading expert in bacteria motility, advocates the indispensability of cognitive concepts for understanding “the knowledge that a cell has of itself”. Perkins and Swain (2009, 12), in turn, claim that “human-developed theories of decision-making under uncertainty apply at the cellular level as well”:

“In situations where the signal fluctuates substantially over time, the cell might be expected to continually update its beliefs. [...] E. coli appears to infer from a sudden increase in temperature that it has left the soil and is now in a host organism. [...] Once a cell has inferred the most probable state of its environment, it needs to decide an appropriate response.” (Perkins and Swain 2009, 4-5)

The assumption that goal-directed agency is the mark of cognition is also well entrenched in naturalistic philosophy, even across very different views of the metaphysics of cognition. For Dretske (1980, 284) for example, goal-directed behavior is constituted as such by its intentional etiology:

“Is it appropriate for the hen not to run from the fox? This depends ... on whether ... she wants to protect her chicks, on what their purposes are. Independently of these factors, the hen’s behavior is neither appropriate nor inappropriate. To describe the hen, for example, as engaging in diversionary tactics (to protect her chicks) is
already to describe her behavior in a way that presupposes an intentional structure for the internal source of that behavior. The appropriateness of response, then, insofar, as this is relevant to what the organism believes and intends, is a property the response acquires only in virtue of its production by internal states having content.”

Similarly, for Sterelny (1990, 21) goal-directed behavior requires representation and inferential capacities:

“The there can be no flexible and adaptive response to the world without representation. To learn about the world, and to use what we learn to act in new ways, we must be able to represent the world, our goals, and options. Furthermore we must make appropriate inferences from those representations.”

And according to Dennett (2009, 3) the agency of a system emerges only from the intentional stance. This stance, we are told, is

“a subspecies of the design stance, in which the designed thing is treated as an agent of sorts, with beliefs and desires and enough rationality to do what it ought to do given those beliefs and desires”.

So to interpret a system as an agent is to interpret it as the locus of cognitive attribution. As Dennett puts it, “the goals of goal-directed computer must be specified intentionally” (1971, 91).

A central motivation for this assumption is the fact that cognitive agency, the capacity to act for a reason, is the paradigm of goal-directed agency. As we saw in Chapter 1, we typically explain why a cognitive agent acted in a particular way by citing her intentions or desires, which represent her goals, and her instrumental beliefs, which represent what she takes to be the means to attain her goals. These explanations are rationalizing (Davidson 1963). They explain by making sense of the behavior as rationally justified (or not) in light of the agent’s intentions and the norms of reason. As such, these explanations have normative implications about what the agent ought to do given her intentions and circumstances. Cognitive agency thus motivates two arguments for the conclusion that goal-directed behavior is the mark of cognition, the argument from intentionality and the argument from normativity. These arguments were introduced in
Chapter 1 as standard objections against natural teleology. I will present them again to motivate the premise that agency presupposes cognition.\(^\text{30}\)

A teleological explanation explains the occurrence of some event by showing it to be the (or a) means to the attainment of some goal \(G\). But by the time the means occurs, \(G\) is an unactualized future state of affairs. So how can an unactualized future state of affairs explain the occurrence of the means? Cognitive agency provides a model that avoids commitment to backwards causation. It is not the unactualized future state of affairs per se that explains the occurrence of the means. It is the representation by the system of that state of affairs which causes and hence explains the occurrence of the means. Goal-directed action thus requires intentionality, the mental representation of goals by the intentions of the agent.

Now, if an agent has goal \(G\) and doing \(r\) is the means to attain \(G\) under circumstances \(C\), then ceteris paribus the agent ought to do \(r\). But for the means to be something that the agent ought to produce, \(G\), some argue, must be a state of affairs that ought to be attained (Bedau 1992). But for \(G\) to be a state of affairs that ought to be attained \(G\) must be good. To avoid the worry of allowing intrinsically evaluable states of affairs in our ontology, cognitive agency again provides a naturalistic model. It is not \(G\) per se that is good but the representation of \(G\) as good or under “the guise of the good” hence as something to be pursued (Velleman 1992). Goal-directed behavior thus requires rationality, the capacity to respond to the norms of reason.

The conclusion that bacteria are cognitive agents follows then from two premises: That bacteria are goal-directed agents; and that goal-directed agency is the mark of cognition. The former is an empirically observable fact about the characteristic robust adaptive dynamics of living systems. The latter seems to follow on naturalistic grounds from considerations about the nature and explanatory role of goal-directed capacities. But we have seen in the previous chapter that there is an alternative account of natural teleology—the invariance account (Walsh 2012)—that is not committed to backwards causation or intentionality and that coupled to an ecological account of agency vindicates a naturalistically acceptable form of normativity.

\(^{30}\) These are adapted from Walsh (2008).
5.2.3. The Role of Cognition in Microbiology

To be sure, the claim is not that bacteria have human level cognitive capacities but that the fundamental elements of cognition are present in bacteria. Still, the role that cognition is supposed to play in this research can be interpreted in different ways, not all of which are committed to realism about unicellular cognition. At worst, attributing cognition to unicellular organisms can be interpreted as an anthropocentric projection of our own cognitive agency onto the natural world. As such, it plays a distorting role in our scientific understanding of cellular behavior. But the progressive character of this research, namely, the fact that new capacities and new ways to think and model these capacities are being discovered and analyzed under the cognitive hypothesis, strongly speaks against this interpretation.

Another interpretation is that even if attributing cognition to cells is not a pernicious anthropomorphism it is nevertheless a mere metaphor. As such, it plays no serious theoretical role in our scientific understanding of cellular dynamics. Rather, it is a colorful honorific title meant to highlight the surprisingly resourceful character of cell behavior. At best, this metaphor functions as a heuristic tool that guides research until a purely mechanical non-cognitive account can be given. Surely there is plenty of metaphoric use of cognitive language in biology and science more generally. However, the seemingly literal use of cognitive attributions such as “once a cell has inferred the most probable state of its environment, it needs to decide an appropriate response” (Perkins and Swain 2009, pp. 4-5) speaks against this facile figurative interpretation. The mark of theoretical role is explanatory role. Cognitive states are attributed here in order to explain a distinctive pattern of behavior that exhibits the distinctive modal profile of hypothetical invariance.

But on a richer interpretation along these lines one can argue that treating cells as if they were rational agents making inferences and acting on their beliefs and desires is not a mere heuristic device but an effective predictive strategy (Dennett 1987). The idea is that absent a detailed mechanical account, attributing cognition allows us to gain a good grasp on the behavior of these complex adaptive systems. Furthermore, since the cognitive strategy works in practice for the most part, there is no longer a motivation to ask the further theoretical question as to whether bacteria are really cognitive agents. The cognitive strategy is after all only provisional: All “cognitive loans” must ultimately be paid with non-cognitive strictly causal-mechanical
currency, which gives us much safer predictions. So even if cognitive attribution is pragmatically useful as a predictive strategy, it is ultimately theoretically dispensable as an explanation of the behavior of these systems. This pragmatic interpretation has the advantage of allowing us to take bacterial cognition more seriously than the figurative interpretation, but not too seriously.

However, one problem with this interpretation is that we already know a great deal about the signal-transduction mechanisms that underpin cellular behavior. So the role of cognition cannot be provisional pending on such knowledge. But more importantly, the fact that cognitive attribution persists despite mechanical knowledge of cellular behavior indicates that this interpretation fails to capture the distinctive explanatory role that cognitive attribution plays in general and in this research particularly. The mark of indispensability is explanatory power. As we have seen, cognition plays a distinctive role in the explanation of goal-directed behavior. This holds for human as well as for cellular behavior. So perhaps we need to take the role of cognitive attribution theoretically more seriously.

Cognition is attributed in order to explain the behavioral dynamics of cells irrespective of the biochemical details of its mechanical implementation. As in the folk-psychological case, cognitive attribution explains why, that is, for what reason or purpose an agent acted in a particular way irrespective of how or by which neurophysiological mechanisms the behavior was causally produced. The higher-level regularities that cognitive concepts are meant to capture are robust across changes in the underlying causal mechanical details of its implementation and hence are insensitive to those details. Cognitive concepts hence cannot be ultimately eliminated in favor of non-cognitive strictly causal-mechanical concepts without loss of theoretically crucial (modal) information about the behavior of the system. After all, all behavioral responses, whether of bacteria or humans are mechanically caused. Yet, citing the particular neural or signal transduction pathway that caused a particular response fails to distinguish whether the response was randomly produced, hence not purposive, or the result of an “autonomous decision”. To capture the fact that a response was due to an autonomous decision rather than to a stochastic event we need to employ cognitive concepts. Cognitive concepts are thus not just pragmatically useful heuristic or predictive devices. They are theoretically indispensable to explain the behavior of rational agents as rational agents, whether human or bacteria. On account of this, we have strong abductive reasons to think that bacteria really are cognitive agents.
I don’t mean to endorse this argument for a realist interpretation of bacterial cognition. The point of this last subsection is rather to give a rationale for the tendency to take cognitive attributions theoretically seriously against the facile dismissive objection of the metaphoric interpretation. What these arguments show us is that teleology is indispensable to the explanation of the behavior of bacteria.

5.3. Cartesian Agency

5.3.1. The Mechanistic Turn and the Primacy of Cognition

Ironically, the premise that goal-directed agency is the mark of cognition was also central for the mechanistic turn in biology led by Descartes in the 16th century. Descartes sought to oppose the Aristotelian doctrine that the nature of organisms understood as their goal-directed capacities to pursue their ways of life (“bios”) (Lennox 2009) grounded a hierarchy of “souls” or forms:31 Plants have a vegetative form, animals have a sensitive form and human beings have a cognitive (rational) form (Grene and Depew 2004). This framework allowed for different kinds of goal-directed capacities along a continuum, from less sophisticated to more sophisticated, where only highly sophisticated goal-directed capacities familiar from human behavior involved cognition. Phenomena such as growth, for example, constituted an instance of a goal-directed capacity that was not mediated by the intentional representation of the end-state to be achieved. Yet, growth was conceived to be qualitatively distinct from the mere mechanical change characteristic of inanimate matter. Aristotle (1970, 8) thus denied the Cartesian premise that all goal-directed capacities imply cognition. As he put it:

“It is absurd to suppose that purpose is not present because we do not observe the agent deliberating”.

A couple of millennia later Descartes (1639, CSMK III, 134) took a radical turn that denied any uniqueness to the capacities of living beings:32

31 For Aristotle the nature of a phenomenon is an inner principle of change that explains its place in the natural world.

“The number and the orderly arrangement of the nerves, veins, bones and other parts of an animal do not show that nature is insufficient to form them, provided you suppose that in everything nature acts in accordance with the laws of mechanics.”

On this mechanistic view, biological phenomena such as birth, growth and reproduction are just the effect of the local motion between bodily parts, not the manifestation of a vegetative or sensitive goal-directed agency. Furthermore, Descartes held that cognition took place in a distinct substance ontologically independent from the physical world and unique to human beings. Non-human living creatures were thus reduced to mere machines or automata, leaving cognitive agency as the only kind of goal-directed capacity. Goal-directed behavior thus became the mark of cognitive agency.

To support this conclusion Descartes proposed the argument from flexibility (adaptability): Cognition (reason) is the goal-directed capacity to behave appropriately across an unbounded set of conditions. Machines, in contrast, can only operate across a limited set of pre-specified conditions. Since non-human creatures can only operate across a limited set of pre-specified conditions, they are mere machines. Cognition is thus the exclusive commodity of a privileged elite species, human beings. As he put it (1985, 53):

“[A]lthough … machines might execute many things with equal or perhaps greater perfection than any of us, they would, without doubt, fail in certain others from which it could be discovered that they did not act from knowledge, but solely from the disposition of their organs: for while Reason is a universal instrument that is alike available on every occasion, these organs, on the contrary, need a particular arrangement for each particular action; whence it must be morally impossible that there should exist in any machine a diversity of organs sufficient to enable it to act in all the occurrences of life, in the way in which our reason enables us to act.”

Nowadays research in cellular decision-making and bacterial cognition has precisely shown, against this conclusion, that the responsiveness of even the simplest living systems is remarkably

33 In Aristotelian terms, Descartes took the nature of organism to be determined by its matter not its form.
resourceful and flexible. This responsiveness is not *unbounded*. So strictly speaking it does not satisfy the stringent Cartesian condition for cognition. Yet, biologists are taking this responsiveness to be evidence of cognition. In doing so they are implicitly accepting a much weaker condition for cognition, namely, responsiveness across a *wide* range of conditions. Perhaps the motivation is a rejection of the anthropocentric elitism about the scope of cognition that follows from the Cartesian condition. In any case, the important point is that while Descartes denied goal-directed capacities to non-human living systems and hence concluded that they were mere automata, contemporary biologists acknowledge their goal-directed capacities and instead conclude that they are cognitive agents. Both arguments, ironically, are predicated on the Cartesian premise that goal-directed agency is the mark of cognition. This premise thus generates a dilemma between mechanism, the view that bacteria are mere automata, and intellectualism, the view that bacteria are full blown cognitive agents.

### 5.3.2. The Dilemma between Mechanism and Intellectualism

But this dilemma forces a problematic choice. Mechanism, on the one hand, nicely preserves the intuition that there is a fundamental difference between bacteria and full-blown cognitive agents. Cognitive agents are responsive to the norms of reason rather than just to the biological demands of survival and reproduction. This difference marks an important discontinuity between living and cognitive agents. But by making this a difference in *kind*, mechanism fails to preserve the intuition that like cognitive agents, bacteria are purposive entities (agents). This marks an important continuity between living and cognitive agents (Godfrey-Smith 1996; Thompson 2007). On the other hand, intellectualism preserves the intuition that bacteria and humans are alike in pursuing goals and hence vindicates the continuity between living and cognitive systems. But on the other hand, by taking the life-mind continuity to be cognitive, intellectualism fails to support the intuition that humans differ from bacteria in being responsive to rational (non-biological) norms.

Mechanism has been defended on the grounds that language is necessary for having concepts and beliefs and hence for responding to rational norms (Davidson 1982; Brandom 2000). An alternative defense holds that unicellular organisms merely transduce but don’t semantically represent stimuli and hence lack the capacity to respond inferentially to non-nomic properties.
(Fodor 1986). But the problem with mechanism is that it radically underestimates the agency of unicellular organisms. For example, Kirschner et al. (2000, 79) tell us that

“Bacterial chemotaxis also appears superficially to be a simple signal-response machine, where an attractant or repellent is perceived by receptor on the bacterial surface to generate a signal that is converted to directed movement. We could imagine all sorts of linkages that would control a motor or a steering mechanism to guide the bacterium by chemical signals. In fact, bacterial chemotaxis is based on the modulation of random movement by ligand binding, resulting in a biased random walk. The specific path any bacterium takes is not directly informed by the binding of the ligand, nor does the individual bacterium at any moment sense a spatial gradient (Berg, 1988). This is quite different from any machine of human design! Biological systems look even less like machines when one considers they can generate order from disorder and can arrive at functional states and responses over a range of starting points, sized of components, and sized of final product.”

Unlike machines that are the product of design, organisms are embodied, self-producing, self-maintaining, and self-organizing complex adaptive systems that synthesize the very materials out of which they are made (Keller 2007; Walsh 2013). The goals of a machine are determined by the intentions of its designer and hence are extrinsic to the system. But the goals of an organism are determined by the organism itself and hence are intrinsic to the system (Jonas 1966; Nicholson 2013). So mechanism fails to accommodate the fact that living systems are the agents of their own adaptive responsiveness, rather than the mere vehicles for the replication of genetic information. Dawkins (2006, xxi) captures this mechanistic view succinctly:

“We are survival machines – robot vehicles blindly programmed to preserve the selfish molecules known as genes.”

In contrast, Bray (2012, 196) gives voice to the poverty of mechanism to capture “the knowledge that a cell has of itself”:

34 Similarly Brooks (2012, 462) tell us that “biological systems clearly differ [from Turing machines]. They respond to varied stimuli over long periods of time; those responses in turn alter their environment and subsequent stimuli.”
“But what words can we use to describe this cellular information? Contemporary biology embraces reductionism and eschews vitalism. It has been inordinately successful in revealing the structures and function of biological molecules, often at an atomic scale. But it has left us with an extreme, almost puritanical rejection of any account of biological processes that goes beyond physics or chemistry.”

Faced with the poverty of mechanism Bray is forced towards intellectualism:

“To me, the most natural way to describe this behavior is simply to say: yes, the bacterium *knows* about temperature and what it means for its survival.” (196)

But intellectualism faces the converse problem of overestimating bacterial agency. After all, however resourceful and flexible bacterial responses are, they fall short of the unbounded responsiveness that Descartes reserved for full-blown cognitive agents. Unicellular organism can be said to act in pursuit of biological goals. But the adaptive dynamics of bacteria indicates no responsiveness to rational norms and hence no capacity to think or deliberate about how to act. Attributing cognition to these systems thus seems an excessive stretch that blurs theoretically important distinctions. Burge (2009, 256) expresses the excesses of intellectualism:

“Action theory … has been almost as hyper-intellectualized as perception theory. Usually discussion begins with cases involving desire, intention, will … There is nothing in itself wrong with this focus, of course. But often it is assumed that such approaches encompass all action. Animal action begins earlier. Much of it is pre-intentional, even pre-representational.”

So while mechanism fails to capture bacteria as agents, intellectualism succeeds but by overestimating their agency as cognitive. There is no intermediate position in this conceptual scheme for unicellular organisms to occupy. Davidson (1999, 11) articulates this tension:

“We have many vocabularies for describing nature when we regard it as mindless, and we have a mentalistic vocabulary for describing thought and intentional action: what we lack is a way of describing what is in between. This is particularly evident when we speak of the “intentions” and “desires” of simple animals; we have no better way to explain what they do.”
I believe that we can do better. This dilemma is an instance of the dichotomy identified in Chapter 1 between eliminativism and primitivism. To carve out a via media between these two extremes and avoid the Cartesian dilemma I propose to follow Aristotle and reject the premise that goal-directed agency is the mark of cognition. For this purpose I appeal to the ecological theory of agency. But before that let us briefly consider an alternative approach that aims to avoid the Cartesian dilemma.

5.3.3. Minimal Cognition

According to the minimal cognition approach (van Duijn et al. 2006; Lyon 2006; Calvo and Keijzer 2011), the dilemma is a consequence of the narrow anthropocentric conception of cognition as “reasoning” that is standard in cognitive science. The problem, they argue, is that this conception applies only to an elite group of organisms thus leaving the rest as mere automata. The populist solution they propose is to abandon the anthropocentric conception in favor of a much broader biological conception of cognition that can be applied all the way down to the simplest unicellular organisms. Cognition, they argue, is fundamentally sensorimotor coupling. Bacteria are thus not mere automata since they are minimally cognitive. But neither are they full-blown cognitive agents since their cognitive capacities are precisely minimal, they involve no reference to intentional states. Minimal cognition, it seems, allows us to navigate a via media between the extremes of mechanism and intellectualism.

This is a revisionist conception of cognition. The question is whether it can play the distinctive explanatory role that the standard conception plays in ordinary and scientific contexts. Cognition figures in rationalizing explanations of behavior. In particular, cognitive states underwrite rational norms that impose certain demands on what an agent ought to do. These explanations, we have seen, capture behavior as purposive action, as opposed to as mere physical movement. However, explanations in terms of sensorimotor coupling are mechanical. They explain how but

35 Merleau-Ponty’s account of perceptual experience as essentially embodied aims precisely to capture such a via media. As he puts it (1962, 144) “... we are brought to the recognition of something between [mechanical] movement as a third person process and thought as a representation of movement – something which is an anticipation of, or arrival at, the objective and is ensured by the body itself as a motor power, a ‘motor project’, a ‘motor intentionality’.” By “motor intentionality” he means a skillful know-how or bodily understanding of the world that does not involve cognition but that can neither be reduced to mere mechanical movement on the grounds that it involves being directed towards an object.
not *why*, that is, for what reason of purpose, the behavior was produced. Hence, they lack the characteristic modal and normative implications of rationalizing explanations. So it seems that the revised conception cannot capture the distinctive theoretical role that cognition plays in explanation.

It may be that having a sensorimotor organization is a causally necessary and/or sufficient condition for having cognition. That is, it may be that in living systems cognition is partly realized in the causal structure of the world by a sensorimotor organization coupled to the environment. But it doesn’t follow unless we assume D1 that the concept of cognition is the concept of having a certain causal-mechanical organization. Not surprisingly, having identified cognition with a putative condition that causally realizes it in the world, this approach has to specify the theoretical role of the concept of cognition by citing its causal realizer as per D2. So the minimal cognition approach is committed to G3 and hence faces the problem of eliminating the normativity of cognition.36

Sensorimotor coupling is after all a sophisticated instance of adaptive coupling. The problem is that cognition seems qualitatively different from the other adaptive capacities of organisms. After all much cognitive behavior is maladaptive. While cognition is constituted by non-biological rational goals, arguably all adaptive capacities are constituted as such by biological goals. If so, then cognition captures a distinctive set of capacities that play a distinctive theoretical role in understanding (intelligent) behavior. But the minimal cognition approach takes all these distinctive capacities including biological agency and cognition as located on a single dimension such that they can only differ in degree. So the minimal cognition approach risks blurring important theoretical distinctions that are explanatorily indispensable.

Finally, notice that this is an account of cognition, not agency. And although the sensorimotor-coupling conception is an embodied, action-oriented hence anti-intellectualist account of cognition, it is still committed to the Cartesian assumption that agency presupposes cognition.

---

36 Calvo and Keijzer (2011) for example discuss minimal forms of cognition, and consider motility and having a dedicated sensorimotor organization as key features for designating the domain of minimal cognition. They conclude that plants are minimally cognitive. They also argue—to make explicit the mechanistic tendencies—that since plants have a control center for behavior dispersed across the root tips we should talk about “plant neurobiology”.
Thus, instead of bridging the gap between the cognitive and the non-cognitive, the strategy is to expand the domain of cognition by re-drawing the lower boundary of the gap. They claim that their account shows the existence of “a vast cognitive spectrum that fills the gap between the mindful and the mindless.” But strictly speaking a “cognitive spectrum” cannot fill “the gap between the mindful and the mindless” because cognition is of course part of the domain of the mindful. To fill the gap we need something that is neither mindful (cognitive) nor mindless (mere mechanism). This, I argue, is a neo-Aristotelian notion of agency that does not presuppose cognition.

5.4. Ecological Agency

5.4.1. Agency as an Ecological Concept

In the previous chapter I argued that an agent is an ecologically embedded purposive system. That is to say, an agent is any system with the gross behavioral capacity to bias its repertoire in response to what its conditions afford for attaining its goals. This set of ecological concepts—repertoire, affordance, goal—makes a tight, interdefined cluster: A system has a goal only if it can respond to affordances and, reciprocally, a set of physical conditions constitute an affordance only if a system has in its repertoire the capacity to respond to it in a way relevant to its goals. Equally, a repertoire is only a repertoire if it can be biased by the system to produce a response that is appropriate for attaining its goals given what its conditions afford.

Agency, in this view, is grounded in the (G2) natures of organisms understood as their goal-directed capacities to pursue their ways of life, where an organism’s way of life is constituted by the set of affordances that it encounters and hence by its goals and repertoire. The distinctive Aristotelian character of this account of agency has two crucial implications that set it apart from the Cartesian approach. First, it implies that agency cannot be given a strictly causal-mechanical definition (G3). Agency is rather defined in terms of the purpose of natural purposive systems (organisms). Second, it entails that agency does not presuppose cognition. Let us now consider how, free from these Cartesian assumptions, the ecological account of agency allows us to overcome the mechanism vs. intellectualism dilemma in favor of a neo-Aristotelian continuum of purposive agency that vindicates a via media for unicellular organisms.
5.4.2. A Continuum of Natural Agents

An agent is any system that can respond to affordances. But the repertoire of an agent can vary in its richness along a continuum, from the simplest biological agents such as unicellular organisms, to the most complex cognitive agents such as human beings. The broader the repertoire of a system, the broader the array of affordances it can respond to and hence the wider the range of goals it can pursue. As we move along the continuum, organisms become increasingly more capable of responding to affordances until the complexity of their ecological dynamics makes belief-desire psychology an indispensable tool for explaining their supple adaptive dynamics (more on this below). The ecological approach thus allows for degrees of agency across a broad spectrum of behavioral profiles. Cognitive agency is thus not agency tout court but only a highly sophisticated instance.

The minimal cognition approach also conceives the difference between bacteria and human beings as a difference in degree along a continuum. But there are important differences. First, on the ecological approach there is a continuum of agency, not cognition. This is important since the dilemma between mechanism and intellectualism in generated by the assumption that agency depends on cognition rather than the other way around. Second, the minimal cognition approach takes a one-dimensional view of the continuum. The idea is that each new element in what I call the repertoire of an agent is just the amplification or sophistication of a previous one, its precursor. But on the multidimensional view of the ecological approach there can be differences of kind along the continuum. As the agency of organism’s increases in its richness, it takes on qualitatively new elements to its repertoire, for example, cognition. Surely there are degrees of cognition. But it doesn’t follow that there are precursors of cognition in previous forms of non-cognitive adaptive agency. For example, as the metabolic rates of organisms increase, new elements of the repertoire can emerge, for example, new modes of locomotion may become possible, such as walking and flight. These don’t have precursors in simpler forms of agency. They are qualitatively new elements of the repertoire.

Unicellular organisms have comparatively the most limited repertoires. In chemotaxis there are only two modes of locomotion, swimming in a straight line or tumbling at random. Because of their limited repertoire, bacteria can respond only to a comparatively limited range of chemical affordances and hence they can attain their goals only across a very limited set of conditions.
Furthermore, the range of goals that they can pursue is limited to the basic biological demands of their life-cycle. Still, their activities are supple and adaptive enough to warrant treating them as responses to affordances, that is, as hypothetically invariant, hence as agential. This adaptive agency is surely minimal, but is agency nonetheless. And crucially, it does not presuppose cognitive capacities, however minimal, nor the teleological mode of explanation they afford requires the conceptual apparatus of intentional psychology.

As we move along the continuum new repertoires emerge and hence new and more complex forms of agency appear. The basic idea is that the emergence of new mechanisms expand what a system can do and hence enable a new repertoire. This new repertoires make for the possibility of new affordances, which in turn allows for a new range of goals.

For example, consider a paramecium. Unlike bacteria, which are unicellular prokaryotes, paramecia are unicellular eukaryotes that prey on bacteria. Yet, despite their more complex internal organization, their motile behavioral repertoire is not that different from the repertoire that bacteria exhibit in chemotaxis. Paramecia spiral through water as they advance. They propel themselves by means of the whiplash movement of their cilia. For each beat of its cilium there are two phases, a fast “effective stroke” and a slow “recovery stroke”. When they encounter an obstacle the effective stroke is reversed in order to swim backwards for a period of time before it resumes it forward movement. These two modes of locomotion are biased by the paramecium in response to what its conditions afford for fulfilling its life cycle. Despite the very limited repertoire of these unicellular organisms, the supple adaptive dynamics that is thus achieved and maintained exhibits the characteristic hypothetically invariant profile of purposive agency. Hence, we can specify its goals by citing its repertoire and affordances. These goals, furthermore, allows us, to say that the paramecium is swimming backwards in order to avoid an obstacle and to explain why it responds in the way it does. This explanation makes no reference to intentional states nor presupposes rational capacities. So the ecological approach allows us to vindicate the agency of paramecia while avoiding the poverty of mechanism and the excesses of intellectualism. But notice that its repertoire, affordances and goals are not that different from the case of the bacterium.

In contrast, consider another single-celled organism, the slime mold. Recent experimental evidence shows that slime molds tend to avoid retracing their own paths. This has motivated the
hypothesis that they use “externalized spatial memory” to navigate their surroundings (Reid et al. 2012). By leaving behind a trail of slime in its path the slime mold can later detect where it has already been. They can also recognize and respond to the trails left by other species of slime molds. These unicellular organisms are also known to solve mazes and anticipate periodic events (Nakagaki et al. 2000). Slime molds possess different mechanisms than those of bacteria or paramecia. The mechanisms expand what they can do and hence enable a new repertoire. For example, the repertoire of the slime mold includes distinctive ways to navigate its surroundings that are unavailable to bacteria or paramecia, such as moving in various directions at the same time. In turn, this new repertoire makes for the possibility of new affordances, such as the possibility to detect where it has been before and hence to respond appropriately. And with new affordances, the system can pursue a whole new range of goals. For example, unlike the bacterium or the paramecium, a slime mold can aim to follow or avoid the trails left by other organisms. As per the ecological theory of agency, we can specify what its goals are in terms of its repertoire and affordances. And it is this fact that constitutes slime molds as adaptive agents.

Finally, consider the squirrel in the park mentioned in the previous chapter. Its repertoire is comparatively much richer than that of bacteria, paramecia and slime molds. In particular, the motile capacities of the squirrel provide it with a set of responses that allows it to do a wider variety of things well beyond moving in a straight line or tumbling at random. The squirrel can jump, run, stay still and modulate its velocity appropriately, among other responses. Because of this rich repertoire the affordance landscape of the squirrel as well as the set of goals it can pursue are also substantially much richer that in the case of unicellular organisms. Yet, arguably the ecological profile of the squirrel is, like that of unicellular organisms, still limited to the biological demands of survival and reproduction. On account of this it would seem the conceptual apparatus of intentional psychology is not necessary to explain teleologically the behavior of the squirrel in the park.

As these examples show, ecological approach avoids the two extremes of mechanism and intellectualism. But it does something subtler than this. The ecological approach transforms the problematic Cartesian dilemma into an Aristotelian continuum with mechanism and

\[37\] For an account of the possible changes in the internal organization and dynamics of organisms that underpins the emergence of new forms of adaptive agency see Moreno and Mossio (2015).
intellectualism at each of its poles. On this graded picture bacteria and paramecium are very close to the mechanistic pole while the squirrel is very close to the intellectualist pole and the slime mold is a more intermediate case. In fact, since the ecological approach is an instance of G2 there is no problem in accepting borderline cases. For example, arguably bacteria constitute such a case: automata approaching minimal agency. In the same way, presumably the squirrel constitutes another borderline case: adaptive agents approaching cognitive agency. In any case the point is that the via media that this continuum allows makes room for a potential measure of the distance from each of the extreme poles. On the basis of this measure along the continuum we can in principle locate the agency of an organism.

Repertoires can be more or less rich. But it doesn’t follow that there is no discontinuity in capacities or in the specific goals. The continuum of agency is consistent with the emergence of novel repertoires and hence qualitatively distinct forms of agency, including cognitive agency. The ecological approach takes cognition to be one such discontinuity. Cognition is qualitatively different, as Descartes argued, from the other adaptive capacities of organisms. While cognition is characterized by non-biological rational goals, arguably all adaptive capacities are constituted as such by biological goals. If so, then cognition captures a distinctive set of capacities that play a distinctive theoretical role in understanding (intelligent) behavior. This role is teleological but it is a distinctive kind of teleological explanation, namely, one that appeals to the conceptual apparatus of intentional psychology.

As Descartes observed, cognitive agents possess an almost unbounded repertoire that allows them to respond appropriately across an almost unlimited range of affordances. This highly complex behavioral profile allows them to pursue an almost unbounded range of goals well beyond those imposed by mere biological demands. In particular, their rich repertoire and affordances landscape allows them to pursue cognitive and conative goals and hence to respond to the norms of reason. Unlike mere adaptive agents then, cognitive agents can pursue goals that go beyond the biological demands of survival and reproduction. In this sense the extremely reach repertoire of cognitive agents discloses a whole new affordance landscape that makes available a novel normative dimension for evaluating purposive behavior.

The highly complex supple adaptive dynamics of cognitive agents requires the conceptual apparatus of intentional psychology. We attribute desires and intentions to cognitive agents
irrespective of whether their satisfaction promotes or impedes their biological well-being. We also explain their behavior in terms these intentions and judge them to be rationally justified or not according to whether they contribute to the fulfillment of those goals. Intentional attributions are hence made true not by discrete facts about the inner workings of the agent but by relational facts about the overall dynamics of the whole ecological system. Cognitive agency is thus a highly sophisticated kind of natural agency but still continuous with the minimal agency of unicellular organisms.

The ecological approach thus provides us with the via media necessary to resolve the tension we encountered in attributing cognition to single cells and hence to occupy the “in between” space that Davidson thought was unavailable. The fact that bacteria have a behavioral profile that can be described ecologically shows, against mechanism, that they are not automata but adaptive agents: They are responsive to the relevance of their conditions for attaining their biological goals. But the fact that their profile is very limited shows, against intellectualism, that bacteria are not cognitive agents. In contrast, the Cartesian framework offers us a dichotomy between mechanism, which entails the complete lack of agency, and intellectualism, which entails full-blown cognitive agency, with no gradations in between. But on the ecological approach mechanism and intellectualism occupy either pole of a spectrum. As we move along the continuum, agential systems are increasingly more capable of responding to affordances until the conceptual apparatus of intentional psychology becomes indispensable.

5.5. Conclusion

I have argued that the cognitive turn in microbiology and cell biology is predicated on a traditional Cartesian conception of agency according to which goal-directed behavior is the mark of cognition. This conception, however, imposes a dilemma between mechanism, the view that unicellular organisms are mere automata, and intellectualism, the view that their actions exhibit full-blown rationality. But while mechanism fails to capture bacteria as agents, intellectualism overestimates their agency as implying full-blown belief-desire psychology. Unicellular organisms, however, occupy a middle way between these two extremes. To carve out this in between space I have used the ecological conception of agency as the capacity to respond to conditions as affordances. This account, I argued, allows for degrees of agency along a
continuum, from mere biological agents such as unicellular organisms to the most cognitively sophisticated ones such as human beings. As we move along the continuum, organisms become increasingly more capable of responding to affordances until the complexity of their ecological dynamics makes belief-desire psychology obligatory. This conclusion, I hope, vindicates the importance that these empirical discoveries have for our understanding the nature of agency while avoiding its Cartesian overtones.
Chapter 6
The Case of Action Explanation

6.1. Introduction

This chapter brings together all the work done in the dissertation. It shows that the dissertation identifies a problem and provides a solution that can be generalized to other normative phenomena. One of the central problems in philosophy is the perceived impasse between naturalism and normativity (see De Caro & Macarthur 2008; 2010). My speculation is that the dichotomy between eliminativism and primitivism pervades this problem across many areas in philosophy. Here I make the case for a very specific instance, namely, the problem of whether reasons that explain actions are causes or norms and hence whether action explanations are really causal explanations or irreducibly teleological.

What is it to explain an action? We typically explain why an action occurs by citing the agent’s reason for doing what she did. An action is thus something done by an agent for a reason. An adequate account of what it is to explain an action must thus identify both the causal role that the agent plays in bringing the action about—the fact that actions are done—and the normative role that the agent’s reason for acting plays in justifying the action—the fact that actions are done for a reason. In short, it must locate reasons in the natural-causal order of the world and yet preserve their distinctive normative role in the explanation of action. However, each of the two standard accounts, the causal and the teleological theory of action, vindicate one aspect at the expense of the other, but neither provides a complete unifying account of the causal and normative role that reasons play in the explanation of action. This tension generates a familiar dilemma between eliminativism and primitivism about the normative role of reason in action-explanation. I propose as a diagnosis that while eliminativism is predicated on the assumption that reasons are individuated exclusively by their causal-role, primitivism is predicated on the assumption that reasons are individuated exclusively by their normative-role.

To avoid this dilemma, in section 6.4 I offer in outline an alternative approach, the ecological theory of action, according to which reasons are individuated by the ecological-role an action
plays within an agent’s system of goals, affordances and repertoire. An action’s ecological role is the contribution it makes to the attainment of the agent’s goals (given the agent’s affordances and repertoire). This approach, as sketchy and programmatic as it is, allows us to understand action as being both caused by agents and justified by the agent’s reasons. The aim of this alternative ecological approach is not to conclusively refute the standard accounts. My aim is to gesture to a way in which the ecological theory of agency together with the invariance account of teleology can be applied to a longstanding problem in the philosophy of action. It is notable that recent collections in the philosophy of action (e.g. Aguilar & Buckareff 2010; D’Oro & Constantine 2013; Constantine 2009) include papers that mostly keep developing and finessing the standard accounts but include almost no novel approaches that seek to go beyond the reasons vs. causes impasse.

6.2. The Problem of Action Explanation

Suppose that Lucy stole a car and we want to understand why she did it. Knowing that Lucy stole the car in order to escape from the police provides such an explanation. This explanation is teleological. It has the form ‘A did φ in order to ψ’. The idea is that Lucy (A) stole the car (φ) because she aimed to escape from the police (ψ) and given her circumstances stealing the car was the best means to attain her goal (ψ). The fact that stealing the car was the best means to escape from the police is the reason why she did it. Action-explanations are thus rationalizations; they explain the occurrence of an action by describing it as something done by an agent for a reason.

But what it is for an agent to do something for a reason? To say that stealing the car is something that Lucy did implies that Lucy caused the car to be stolen. An adequate account of action-explanation must thus identify the cause of the action and in particular the agent’s role in causing the action. This causal character of action, in turn, implies that actions are part of the natural-causal order of the world. So an adequate account of action-explanation must be naturalistic. Call

---

38 Action explanations might take an explicit psychological form: A did φ because A desires or intends to ψ and A believes that Doing φ is the means to ψ. A’s goal ψ is the content of her conative state or pro-attitude while the means φ is the content of the instrumental belief.

39 For something to favor or make an action valuable it is not necessary that the agent believes that doing it would be good or that the agent wants to do it. Reasons can be identified with what is believed or desired, not with believing and desiring itself (Dancy 2000; Thompson 2008).
this desideratum naturalism.

Now, Lucy caused the car to be stolen not just accidentally but for a reason, namely, because given her circumstances stealing the car is the best means to escape from the police. In light of the (instrumental) value that stealing the car had for escaping from the police Lucy is thus justified in acting in that way. An adequate account of action-explanation must thus identify the reason for the occurrence of an action and in particular the reason’s normative role in justifying the action. Action-explanations thus have a distinctive normative (rationalizing) content. So an adequate account of action-explanation must also be normatively acceptable. I will call this desideratum normativity.

A complete, unifying account of action explanation must thus satisfy two desiderata, naturalism and normativity, each of which represents an essential intuition about what it is to explain an action in terms of the agent’s reasons for bringing that action about. The problem of action-explanation as I will understand it is the apparent impossibility of vindicating both intuitions and hence of satisfying both desiderata at the same time. We have seen this problem before. It is simply a version of the problem of naturalizing teleology that I discussed in Chapter 1. In the next section I introduce the standard causal and teleological theories of action explanation and argue that they also face the dilemma between eliminativism and primitivism.

6.3. The Standard Accounts and Their Discontents

6.3.1. The Causal Theory of Action

The causal theory of action (CTA) is the view that action-explanations are really just psychological instances of causal explanations (Davidson 1980; Dretske 1988; Goldman 1970):

\[
\text{CTA} = \text{def: } A \text{ did } \phi \text{ in order to } \psi \text{ if and only if } A \text{ did } \phi \text{ because } A \text{ intends to } \psi \text{ and } A \text{'s intention caused } A \text{ (in the right way) to } \phi.
\]

40 Reasons justify actions in the minimal (“anemic”) normative sense of being (hypothetically) good or valuable for the agent’s goal or purpose. As Davidson (1963, 9) put it: “There is an irreducible—though somewhat anemic—sense in which every rationalization justifies: From the agent’s point of view there was, when he acted, something to be said for the action.”
According to CTA then, Lucy stole the car in order to escape from the police if and only if Lucy had the intention of stealing the car and her intention caused her (in the right way) to steal the car.

The central argument for CTA is one according to which an agent may have a reason that justifies her in performing an action and yet that reason fails to explain the action unless that very reason *causes* the action (Davidson 1963). For example, suppose that Lucy has a gunshot wound and stealing the car is the fastest way to get to the hospital. Assuming that Lucy intends to get medical attention she has a reason to steal the car, namely, that stealing the car would bring about that she gets medical attention. Or suppose that Lucy is obsessed with cars and she has a strong desire to possess this particular model but she lacks the money to buy it. This also gives Lucy a reason to steal the car. Any of these reasons would justify her in doing what she did. But as it turns out none of these reasons is the actual reason for which she did it so none of these reasons explains her action. Justification is thus not sufficient for explanation. What identifies the intention of escaping from the police as the reason for which Lucy actually stole the car is the fact that this intention caused her to perform the action. The argument thus draws a crucial distinction between a justificatory reason for performing an action and the reason that explains the action, where the latter is identified as the cause of the action. So all explanatory reasons are justifications but not vice versa.

CTA has clear advantages. By identifying reasons with causes CTA vindicates the intuition that actions are caused. Also, by conceiving the relation between reasons and action when reasons explain actions as a causal relation CTA gives a clear meaning to the explanatory force of the ‘because’ in ‘Lucy stole the car because she aims to escape from the police’. As Davidson (1963: 10) puts it,

> “Cause and effect form the sort of pattern that explains the effect, in a sense of ‘explain’ that we understand as well as any. If reason and action illustrate a different pattern of explanation, that pattern must be identified.”

---

41 The invariance account of teleological explanation precisely identifies this alternative pattern of explanation.
By the same token CTA offers a naturalistic account of action-explanation and hence satisfies the naturalistic desideratum. The idea is that since doing something for a reason just is a familiar instance of one (mental) event causing another (behavioral) event, rational agency is an unproblematic part of the natural-causal order of the world. In this sense one might think that CTA unifies our everyday normative understanding of ourselves as rational agents with our scientific understanding of ourselves as physical systems governed by causal-mechanisms (Bishop 1989; 2010). As Davidson (1980, xv) puts it:

“The thesis is that the ordinary notion of cause which enters into scientific or commonsense accounts of non-psychological affairs is essential also to the understanding of what it is to act with a reason … Cause is the cement of the universe; the concept of cause is what holds together our picture of the universe, a picture that would otherwise disintegrate into a diptych of the mental and the physical.”

But CTA also has important disadvantages. First, by identifying reasons with causes CTA fails to vindicate the intuition that actions are done for a reason in the normative sense in which reasons justify actions. The problem of “deviant-causal-chains” captures this difficulty (Davidson 1980 [1973]). The idea is that an intention may cause an agent to act in a way that leads to the satisfaction of the intention and yet that intention provides no rational justification for the action. It thus does not explain the action in the right way. For example, suppose that Lucy intends to give herself up to the police by staying put rather than to escape from them. And suppose that her having this intention agitates her so much that she is paralyzed out of nervousness thus letting the police catch her. Lucy’s behavior was caused by her intention, but her having the relevant intention is not the reason why she did. She did it out of nervousness. So the intention provides no rational justification as to why she did what she did. CTA thus fails to satisfy normativity.

---

42 See Hornsby (1997) for a version of CTA that is not committed to an ontology of events as the relata of causal relations.

43 Although I will not argue for this claim here, these problem seems to be endemic to most causal accounts in philosophy as indicated in Chapter 2—particularly those that require a “in the right way clause”—on account of which causal analysis of normative concepts lead to eliminativism.

44 Some proposed solutions include Aguilar (2010); Bishop (1989); Mele (2003); Peacocke (1979).
These cases are meant to be counterexamples to CTA showing that causation is not sufficient for justification, hence the proviso “in the right way” in the definition of CTA. CTA may be right that justification is not enough for explanation, but it is necessary.

Another disadvantage is that CTA fails to accommodate the distinctive role that the agent as an agent plays in the explanation of the action. Consider the problem of the “absent or disappearing agent” (Chisholm 1976; Melden 1961; Nagel 1986; Velleman 1992). According to CTA a behavior constitutes an action by virtue of its mental etiology. Mental states or events, however, are sub-agential parts and hence not identical to the agent as a whole. So the agent as an agent plays no distinctive role in the explanation of action. Rather, the agent is the mere passive site for one event to cause another but it doesn’t do anything. So even if CTA vindicates the intuition that actions are caused it fails to vindicate the intuition that actions are caused by agents. But the agent’s role in causing an action is required to make sense of the constitutive fact that actions are done rather than merely caused.45

I conclude that CTA satisfies naturalism (causation) at the expense of normativity (justification) and hence gives an incomplete account of action-explanation. Davidson’s alleged unification of our everyday normative conception and the scientific conception is hence only apparent: Identifying reasons with causes captures neither the normative role that reasons play as reasons in justifying an action nor the causal role that agents as agents play in bringing the action about.

6.3.2. The (Anti-Causal) Teleological Theory of Action

The teleological theory of action (TTA) holds that action-explanations are what they seem to be on the surface, namely, teleological explanations (McLaughlin 2012; Melden 1961; Ginet 1990; O’Connor 2000; Schueler 2003; Sehon 2005; Taylor 1964; Thompson 2008; Wilson 1989). Instead of identifying the causal antecedents of an action as physiological explanations do, action-explanations identify the goal or purpose for which they are done, as indicated by the telic-connective “in order to”. The basic idea is that by revealing the agent’s goal or intention action-explanations explain by rendering the action intelligible. To say that ‘A did φ in order to

45 An anti-naturalist attempt to vindicate the causal role of the agent as such is to postulate an ontologically distinctive form of causation call “agent-causation” over and above event-causation (Chisholm 1976; O’Connor 2000).
\( \psi \) is to say that A’s doing \( \phi \) was directed at \( \psi \) and A’s \( \phi \)-ing towards \( \psi \) is part of a rational interpretation of A’s behavior.

Teleological explanations are irreducibly normative on the grounds that normativity plays an indispensable role in the interpretation of the behavior of agents and interpretation is constituted by the norm of rationality (Davidson 1963). Interpretation consists in re-describing the action as fitting into an overall pattern of rationality that makes it intelligible, for example, by revealing the agent’s intentions (Anscombe 1959). \(^{46}\) To explain an action is to show that the agent was rationally required to perform it in light of her goals and circumstances (Sehon 2005). For example, to say that Lucy stole the car in order to escape from the police is to make sense of her behavior as conforming to the norm of instrumental rationality. Having a goal or intention does not consist in the possession of causally efficacious mental states or events. Rather, it is a primitive feature of the world and hence cannot be analyzed in non-normative scientific-natural terms.

According to the infamous logical-connection argument for TTA (Anscombe 1959; Hampshire 1959; Melden 1961), action-explanations can’t be causal on account of their distinctive normative character. The idea is that it is of the nature of an action that it is done for a reason. For example, to say that an agent acts (intentionally) is equivalent to saying that a certain normative sense of the question ‘why?’ applies to the agent’s action (Anscombe 1959). So the connection between reasons and actions when reasons explain actions is constitutive or analytic and hence conceptually necessary. However, causes are logically distinct from their effects. So the connection between causes and effects is thus contingent and hence only knowable empirically. Therefore, action-explanations are constitutive not causal in character. Reasons don’t cause actions; they imply (prescribe) normative requirements or demands on agents. Justification is an internal relation but causation is an external relation.

There are clear advantages to this account. First, by identifying reasons with norms TTA vindicates the intuition that actions are done for a reason in the normative sense of justifying the

\(^{46}\) Although most teleologists in the philosophy of action are committed to interpretationism in one way or another, not all interpretationists are explicitly teleologists. I use ‘TTA’ as a convenient umbrella name to capture all normativist, anti-causal approaches to action-explanation.
action. On account of this TTA accommodates the distinctive normative content of action-explanations and hence satisfies the normative desideratum. Another advantage is that it vindicates the intuition that agents as agents figure constitutively in the explanation of action. After all TTA makes no reference to sub-agential states and it takes agents as such to be the proper locus of rational attribution. To act for a goal or reason is constitutive of intentional agency and teleological explanations capture this fact.

However, TTA has the important disadvantage of failing to vindicate the intuition that actions are caused by agents. TTA is explicitly presented as an anti-causal theory where causal relations don’t figure. Indeed, TTA is part of a tradition that makes a radical separation between reasons and causes (Sellars 1963; McDowell 1994). So even if TTA vindicates a role for the agent as a whole in the explanation of action it fails to vindicate the agent’s distinctive causal role. Notice that while CTA vindicates the intuition that actions are caused but not that they are caused by agents TTA vindicates the intuition that agents play a role as agents in the explanation of action but not a causal role.

Another disadvantage is that since TTA takes the normativity of reasons for action to be a brute fact about the world we can’t explain why normative truths hold by adducing further facts. In particular, we can’t explain what the relation of justification between reasons and action is in strictly natural terms, that is, in terms endorsed by natural science. So TTA cannot explain the natural status of normative facts and hence fails to satisfy the desideratum of naturalism. After all, defenders of TTA argue, the kind of normative understanding that action-explanations provide is different and autonomous from the kind of causal understanding that natural science provides. Normative facts are thus sui generis (Hornsby 1996; McDowell 1994). So provided that TTA makes no commitment to supernatural entities and remains consistent with natural science there is no further demand to locate rational agency in the causal-order and hence no demand to explain normativity in naturalistic terms. As Sehon (2005: 230-231) put it:

“Agents do not fit within that [the natural] order...The realm of reasons, values, agency, beliefs and desires is outside the realm of the natural sciences... That there is a realm of facts that employs concepts having no application to most of the rest of the natural world ... does not mean that we are beyond the physical laws of nature ... We are every bit as much a part of nature as are the inanimate elements that
surround us. But there are also facts about what we value, what we think and, and what we do, and these facts have no counterparts when the subject is a rock or a tree.”

I conclude that by identifying reasons with norms TTA vindicates the normative role that reasons play as reasons in the explanation of action. But it does so at the expense of failing to locate a rational agency in the causal structure of the world (causation). So the fate of TTA is exactly the inverse of CTA. The alleged autonomy of our everyday normative conception from the scientific conception makes a mystery of the causal role that agents play in action-explanation and of their place in the natural order. TTA thus also gives an incomplete account of the nature of action-explanation.

6.3.3. A Dilemma between Eliminativism and Primitivism (again)

Each account captures one important aspect of action at the expense of the other, but neither provides a complete, unifying account of what action-explanation is. This dialectic generates a dilemma between eliminativism and primitivism about the normativity of action-explanation. Eliminativism is the view that reasons account for how actions are caused (naturalism) but says nothing about how they are justified (normativity). CTA does not deny that reasons justify actions; it just has nothing to say about how they might do so. It seems to suggest that justifying an action isn’t part of explaining it. Primitivism, on the other hand, is the view that action-explanations justify actions (normativity) but says nothing about how they are caused or how the justificatory conditions are causally realized (naturalism). Again, TTA doesn’t deny that actions are caused or that they are natural; it just seems to suggest that citing their causes is not part of the explanation of an action. So each view is just silent about the feature of action-explanation that the other captures. Naturalism (causation) and normativity (justification), it seems, pull in opposite direction; they can’t be jointly satisfied.

The standard accounts may resist this dilemma on the grounds that a theoretical account of action-explanation need not vindicate all our pre-theoretic intuitions. As a reductionist account CTA can deny the demand to vindicate the intuitive force of the normativity desideratum by appealing precisely to naturalistic considerations regarding the unification of our everyday normative understanding with natural science. Sacrificing or at least de-emphasizing normativity,
they claim, is a reasonable price to pay for a naturalistic unification. In contrast, TTA can deny
the intuitive force of the naturalistic desideratum by appealing precisely to the autonomy of our
everyday normative understanding from natural science. If the price to pay for preserving the
distinctive normative character of action-explanation is to reject the stringent demands of
naturalism, so be it. But even if it is true that not all pre-theoretic intuitions are to be vindicated, I
take the dialectic between CTA and TTA to show that the normativity and naturalism intuitions
are essential to what it is to explain an action in terms of reasons and hence non-negotiable. So
instead of adding epicycles to the standard accounts we should seek an alternative reconciliatory
account altogether. But before proposing an alternative positive account that satisfies both
desiderata, we need to make explicit the central assumptions generating the dilemma.

6.4. Diagnosis

The dilemma has essentially the same structure as the dichotomy between eliminativism and
primitivism about teleology introduced in Chapter 1. In Chapter 2 we saw that the eliminativism
or primitivism choice about teleology was forced by the presumptive methodology of naturalism
as committed to D1 and D2. In this section I offer a diagnosis for this dilemma in terms of a
distinctive set of assumptions about the individuation and theoretical role of reasons.

6.4.1. Why CTA leads to Eliminativism

Consider the following simple argument for CTA:

(P1) The theoretical role of reasons is to explain actions.

(P2) Action-explanations cite causes.

(C) The theoretical role of reasons is to cause actions.

P1 is justified by common sense. Reasons for actions are fundamental to our folk theory of
agency. Their role is to explain, predict and evaluate why agents as agents act in the way they do.
Any adequate account of action-explanation must vindicate the ordinary fact that we typically
explain why agents act in the way they do by making hypotheses about their reasons for acting.
P2, in turn, is a consequence of the causal character of actions, that is, of their being done or brought about by agents. Actions are occurrences and for occurrences to happen there is a widespread conviction that they must be caused. The explanation of an occurrence typically cites its cause. So the explanation of the occurrence of an action must also cite its cause.

However, the conclusion C doesn’t follow straightforwardly from P1 and P2. This simple argument is an enthymeme. We need to add further premises to derive C. These, I propose, are P3, P4 and P5:

(P1) The theoretical role of reasons is to explain actions.

(P2) Actions explanations cite causes.

(P3) The theoretical role of reasons is exhausted by their individuation conditions.

(P4) Reasons are individuated by their causal role.

(P5) The theoretical role of reasons is exhausted by their role in causing actions.

(C) The theoretical role of reasons is to cause actions.

P3 is an instance of TI the principle according to which the theoretical role of a phenomenon is exhausted by its individuating conditions introduced in Chapter 2. This principle is justified by scientific practice, as we saw in Chapter 3. For example, as a chemical element oxygen is individuated by the atomic structure that causally realizes it. The role that oxygen plays in chemical theory, furthermore, is exhaustively specified in terms of its atomic structure. Temperature, in contrast, is not individuated by the microscopic material constitution that realizes it, since systems with very different material constitutions can instantiate the same temperature. Rather, temperature is individuated by the set of macroscopic behaviors that characterize how the system would behave under certain (thermodynamic) conditions. The role that temperature plays in thermodynamics is thus exhaustively specified in terms of its behavioral profile irrespective of how it is realized. So although the way a natural, scientific phenomenon is individuated can vary depending on the level of organization, its individuation conditions completely specify its role within a scientific theory. The question is how are reasons
for action individuated? CTA’s answer is given by P4. P4 is the conclusion from Davidson’s argument: reasons are causes. Finally P5 = C. Once premises P3, P4 and P5 are made explicit the conclusion C follows.

Having made explicit the assumptions underlying the argument for CTA we can see that it leads to eliminativism by virtue of individuating reasons essentially and exclusively in causal terms, as per P3 and P4. Identifying the theoretical role of reasons with their causal role accommodates the intuition that actions are caused (although not by the agent as agent). Furthermore, it locates reasons in the causal structure of the world and hence provides a naturalistic account. But since causes don’t typically justify their effects, the theoretical role of reasons is then silent about their normative role in justifying actions. Normative relations are in general not causal. So if reasons are essentially and exclusively causes it seems that normative relations are not part of reasons such that action explanations are just causal explanations with no distinctive normative content. For this reason CTA leads to eliminativism about the normative role that an agent’s reasons play in action-explanation.47

6.4.2. Why TTA leads to Primitivism

Consider the following simple argument for TTA:

(Q1) The theoretical role of reasons is to explain actions.

(Q2) Actions explanations cite norms.

------------------------------------------

(C*) The theoretical role of reasons is to justify actions.

Q1 = P1. Q2, in turn, reflects the normative role of reasons in justifying actions. The idea is that action-explanations explain by showing that the action belongs within a pattern that renders the

47 In Chapter 2 CTA was used as an example of a theory committed to dogma D2, the view that the role of a theoretical entity can be completely specified by citing the causal realizer of that role. But it is important to distinguish between versions of CTA that are committed to D2 (Dretske’s and Mele’s) from those that aren’t, such as Davison. In this chapter I’m talking about Davidson’s version of CTA. In this version the theoretical role of reason is a causal role (as opposed to a normative one). But that role is not specified in terms of the causal realizer of that role. So Davidson’s version is not committed to D2.
overall behavior of the agent rational. However, the conclusion C* doesn’t follow
straightforwardly from Q1 and Q2. This simple argument is an enthymeme. We need to add
further premises to derive C*. These are Q3, Q4 and Q5:

(Q1) The theoretical role of reasons is to explain actions.

(Q2) Actions explanations cite norms.

(Q3) The theoretical role of reasons is exhausted by their individuating
conditions.

(Q4) Reasons are individuated by their normative role.

(Q5) The theoretical role of reasons is exhausted by their role in justifying
actions.

--------------------------------------------------------------------
---------------------------

(C*) The theoretical role of reasons is to justify actions.

Q3 = P3. Q4, in turn, is the conclusion of the argument from logical-connection: reasons are
norms (hence not causes). Finally Q5 is practically a restatement of C* that discloses how it
follows from the premises. Once premises Q3, Q4 and Q5 are made explicit C* follows.

We can now see why the second horn of the dilemma, primitivism, is generated, namely by
virtue of TTA’s individuating reasons essentially and exclusively in normative terms, as per Q3
and Q4. Identifying the theoretical role of reasons with its normative role accommodates the
intuition that reasons justify actions. However, norms don’t typically cause the actions they
prescribe. So the theoretical role of reasons is then silent on what causes actions. Causal relations
are in general not part of norms or don’t figure in the specification of a norm. So if reasons are
essentially and exclusively norms it seems that causal relations are not part of reasons. So TTA
cannot capture the distinctive causal role that agents play in the explanation of action. Hence
TTA’s primitivism about the normativity of action-explanation. Furthermore, since the
justificatory relation between reasons and actions cannot be re-described in non-normative
natural terms, TTA says nothing about how reasons are realized in the causal structure of the
world.
We can conclude that the dilemma is generated by the assumption that reasons are individuated either essentially and exclusively causally (as per P2 and P4) or essentially and exclusively normatively (as per Q2 and Q4) but not both. The challenge is then to find an alternative way of individuating reasons that vindicates both, the causal and the normative character of action-explanation. To this end I will use the apparatus of the ecological theory of agency and naturalized teleological explanation introduced in Chapter 4 to offer in outline an alternative approach to understanding action explanation, the Ecological Theory of Action (ETA). But first let us consider a recent attempt to reconcile CTA and TTA.

6.4.3. A Model-Based Compatibilism

According to Menzies (2010) the incompatibility between CTA and TTA is predicated on a shared deductive-nomological account of causal-explanation. He argues that once we abandon this assumption in favor of a model-based approach to explanation we can reconcile the causal and normative character of action-explanation. The basic idea of the model-based approach is that the same idealized model of a rational agent can play two roles: a normative or prescriptive role by virtue of which the model rationalizes the action and a causal or descriptive role by virtue of which the model causally explains an agent’s behavior. So even if normative and causal relations are very different in nature as per the logical-connection argument, they “overlap in extension” (168), at least in successful non-irrational cases. As he puts it,

“On many occasions we deploy a model of a rational agent in its prescriptive role to tell us how a real agent should act in a given set of circumstances and then employ it in its descriptive role to predict how she will act in those circumstances. Models, by virtue of their idealized character have a Janus-faced nature that enables their use in description and prescription.” (166)

Compatibilism requires that “the properties by virtue of which one event rationalizes another are the same properties as those by virtue of which the first event causes the second” (165). This condition is satisfied because the same model is applied to the same real agent, one time in its prescriptive function and the other is its descriptive function. So there is a single (abstract) property or generalization that both justifies and causes the relevant action thus reconciling CTA and TTA.
However, as Menzies recognizes, “no single application of a model to a real system can realize both functions at the same time” (166). Contrary to Menzies action-explanations perform both functions at the same time. The function of explaining why Lucy stole the car in terms of her wanting to escape from the police is not to inform us of two things, namely, that she caused the car to be stolen and, furthermore, that she did it for a reason. Rather, its single function is to inform us that she caused the car to be stolen for the relevant reason. Escaping from the police rationalizes and hence explains stealing the car. This rationalization implies that Lucy caused the car to be stolen (so rationalization implies causal relations). But these are not two different functions. Action-explanations serve one function: rationalize but rationalizations imply causal relations. In fact, it seems wrong to say that action-explanations serve two functions, causing and rationalizing. The view that action-explanations can be decomposed into a causal and a normative function assumes the dichotomy that Menzies is so eager to overcome.48

6.5. Towards an Ecological Theory of Action

Instead of trying to reconcile CTA and TTA I believe we need a different approach altogether. Here I introduce a very sketchy and programmatic alternative I call ‘the ecological theory of action’ (ETA). ETA combines the ecological account of agency and an invariance account of teleological explanation introduced in Chapter 4. Its main thesis is that the theoretical role of reasons is to give the ecological role of an action. Here is the argument:

(R1) The theoretical role of reasons is to explain actions.

(R2) Action-explanations cite reasons.

(R3) The theoretical role of reasons is exhausted by their individuating conditions.

48 To see this consider the question of how reasons are individuated on the model-based approach. Reasons can’t be individuated conjunctively by their normative and by their causal relations for reasons are never rationalizing and causal at the same time. Rather it seems that reasons are individuated normatively as per TTA when the model is used to rationalize and causally as per CTA when it is used to causally explain. So CTA and TTA can be at most successively compatible on the grounds that they constitute different moments of the same generalization rather than two different generalizations.
(R4) Reasons are individuated by their ecological role.

(R5) The theoretical role of reasons is exhausted by their ecological role.

(C**) The theoretical role of reasons is to specify the ecological role of an action.

The central premise is R4 so I will focus on explaining the technical notion of ‘ecological role’. I will argue that individuating reasons by their ecological role accommodates the causal and normative character of action explanation.

The standard strategy in the philosophy of action has been to first give an account of action and then on that basis derived an account of agency as anything that acts. But ETA inverts the order of dependence and begins with an account of agency on the basis of which it derives an account of action and action-explanation. An agent, we saw in Chapter 4, is any system with the gross behavioral tendency to exploit its repertoire in engaging with the network of affordances that promote or impede the attainment of its goals. More precisely, $X$ is an agent just in case $X$ has a goal $G$; $X$ is embedded in a set of conditions $C$ that afford a repertoire of responses $R$; and $X$ has the capacity to implement the response from its repertoire that is appropriate for $G$ given what $C$ affords.

To explain an action, according to ETA, is to identify its ecological role. An action’s ecological role is the role it plays in the agent’s system of goals, affordances, and repertoire.

**Ecological Role** $= \text{def}$: The ecological role of an action is the contribution it makes to the attainment of an agent’s goal given her repertoire and affordances.

The idea is that reasons *conduce* to the attainment of an agent’s goals. The reason why $X$ did $Y$ is that $X$ has goal $G$, and given $X$’s conditions $C$, $Y$ is conducive to $G$ and if $Y$ hadn’t been conducive to $G$ in $C$ $X$ wouldn’t have done $Y$. So the reason why Lucy stole the car is that she has the goal of escaping from the police, and given her setting, stealing the car is conducive to escaping from the police and if stealing the car hadn’t been conducive to escaping from the police Lucy wouldn’t have stolen the car. The notion of ‘conducing’ was introduced in Chapter 4: $Y$ is *conducive* to $G$ in $C$ just in case $Y$ is a reliable way to produce $G$ in $C$ such that $Y$ is good
for $G$. Conducing then is a hybrid Janus-faced notion: simultaneously causal and normative. ‘Bringing about’ is causal notion. But the ecological role of an action also provides the reason for its occurrence: Lucy stole the car for its contribution to escaping from the police. To identify the ecological role of stealing the car is thus to locate the action in Lucy’s system of goals, such as escaping from the police, affordances, such as there being a car nearby, and her capacity to exploit those affordances appropriately, such as the ability to steal a car or use it to escape the police.

For ETA then, the reason for which an agent performed an action is a complex fact about the way an agent is ecologically embedded. In contrast, for CTA the reason is a sub-agential fact about the inner workings of the agent and for TTA it is a primitive fact about its character as agent.

Now we can say how reasons explain action. According to the invariance account of teleological explanation introduced in Chapter 4, to explain why a means $x$ occurred in terms of the goal $y$ that it subserves is to cite an invariance relation according to which, if $y$ were to change to $y^*$ while background conditions $C$ remains constant then some other $x^*$ would have occurred that would had caused $y*$; and if $y$ were held constant while $C$ changes to $C^*$ then some other $x^*$ would have occurred that would have caused $y$. This relation, as we saw, is captured by the notion of ‘conducing’. So to say that ‘Lucy stole the car in order to escape from the police’ means that:

a. Lucy has the goal of escaping from the police.
b. Lucy caused the car to be stolen;
c. Stealing the car is conducive to escaping from the police; and
d. If stealing the car hadn’t been conducive to escaping from the police then (ceteris paribus) Lucy wouldn’t have stolen it.

Consider (a). Having a goal, remember, is gross behavioral capacity of the way Lucy is embedded in its ecological setting. To say what Lucy’s goal is we need to say what her repertoire and affordances are. In turn, (b) implies that stealing the car is in Lucy’s repertoire while (c) implies that producing that item from her repertoire constitutes an appropriate response. Finally, (d) captures the relation of hypothetical invariance between (b) and (a) such that, if under the circumstances stealing had not been a reliable way of attaining the goal, then Lucy would have
done something else that would have attained the goal. And if Lucy had had a different goal the
she would have done whatever is conducive given circumstances to attain that alternative goal.
For example, suppose Lucy has the goal of escaping from the police but instead of a car there is
motorcycle nearby. Then, provided that riding a motorcycle is in Lucy’s repertoire, she would
have stolen the motorcycle. Alternatively, suppose Lucy has the goal of getting medical attention
rather than to escape from the police and that stealing the car is conducive to getting medical
attention. Then Lucy would have stolen the car.

We are now in a position to see how individuating reasons in terms of an action’s ecological role
vindicates both the agent’s role in causing the action and the reason’s role in justifying it. Being
factive, the reason for an action entails that X did Y and hence that X caused Y to occur.\(^{49}\) So
ETA accommodates the intuition that actions are caused by agents and hence the fact that action
are done rather than that they merely happen. Remember that while CTA vindicated the intuition
that actions are caused but not by the agent TTA vindicated the intuition that the agent played a
constitutive role in action but not a causal role. So ETA accommodates the best of both views.
So it is not sub-agential states (parts) that figure in the explanation as bringing about the action, it
is the agent as whole. After all, it is the agent that possesses the repertoire. By accommodating
the constitutive causal role that the agent as an agent plays in the explanation of action, ETA
avoids the problem of the disappearing agent. Furthermore, CTA is right that having a
justification is not sufficient for explanation, causation is required. But against CTA is doesn’t
follow that action-explanations are thereby causal-explanations. Action-explanations are
teleological and mean-ends relations imply causal relations, but they are not mere causal
relations: Causing is necessary but not sufficient for conducing.

In turn, conducing introduces the normative dimension of action-explanation. To say that Y is
conducive to G is to say that Y is appropriate for G or that Y is good for G. X is thus
(instrumentally) justified in causing Y. So to say that stealing the car is conducive to escaping
from the police is to say that stealing the car is appropriate for or is good for escaping from the
police. Lucy is thus justified in causing the car to be stolen. The reason that Lucy had for stealing

\(^{49}\) To say that the reason for which X did Y is factive is to say that if Z was the reason for doing Y then it is true that
X did Y because of Z. A reason isn’t factive if Z is a reason for doing Y, X did Y and it doesn’t follow that X did Y
because of Z.
the car is thus the role that stealing the car has in its ecology i.e. in its system of goals affordances and repertoires. On account this ETA vindicates the intuition that reasons justify actions. TTA is thus right that action-explanations have a distinctive normative content absent from standard forms of causal-mechanical explanation. But TTA is wrong that thereby causal relations are not part of justificatory relations. Purposive-justificatory relations—hypothetical invariance—imply causal relations hence they cannot be primitive.

The debate about the nature of action-explanation in the philosophy action has been framed in terms of an opposition between causal and anti-causal accounts on the assumption that teleology is anti-causal. This anti-causal character partly motivates TTA’s primitivism about the normativity of teleology. But the invariance account of teleological explanation rejects this assumption: A teleological explanation explains the occurrence of the means by showing it to be conducive to the attainment of the goal. To say that means $Y$ is conducive to goal $G$ is to say *inter alia* that $Y$ reliably causes $G$. So although teleological explanations are a distinctive form of non-causal explanation, it doesn’t follow that causal relations don’t figure in teleological explanations. Teleological explanation is thus not anti-causal. Furthermore, given the hybrid, Janus-face character of conduciveness, the distinctive hypothetically invariance modal profile of this causal relation has normative implications: $Y$’s being conducive to $G$ implies that $Y$ is good for $G$. So the normative relation where goals require their means is causally realized. So ETA accommodates the normativity that TTA is so eager to protect without the primitivism that prevents TTA from locating reasons in the natural order.\(^{50}\)

Finally, ETA accommodates Davidson’s a/the-reason distinction without reducing reasons as causes and hence without reducing action-explanations to causal-explanations. Lucy has a reason to steal the car just in case Lucy has the goal of escaping form the police and stealing the car is conducive to escaping from the police given her setting. In turn, what constitutes this reason as the reason for which she did it is that her having done so supports the following counterfactual: if

\(^{50}\) Thus Mele’s (2003, 39) objection against TTA: “In virtue of what is it true that a person acted in pursuit of a particular goal? No proponent of ‘anticausal teleologism’, as far as I know, has offered a plausible, informative answer to this question.” In light of our discussion we can respond that the reason why the answer is not very informative is that for TTA truths about acting for goal are primitive truths, hence cannot be accounted for in terms of something else. In contrast, ETA provides ecological truthmakers for goal-attributions. As per G2, these ecological truthmakers locate goals in the natural order and preserve their distinctive explanatory role.
she hadn’t had the goal of escaping from the police then (ceteris paribus) she wouldn’t have stolen the car. So it is the reason only if the following counterfactual holds: if she hadn’t had the goal then ceteris paribus she wouldn’t have performed the action for that goal. In contrast, merely having a reason or no reason fails to support this counterfactual.

6.6. Conclusion

An action is something done by an agent for a reason. An adequate account of the nature of the explanation of action must vindicate the causal role that the agent plays in causing the action and the normative role that the agent’s reasons play in justifying the action. But while CTA vindicates causation at the expense of justification, TTA vindicates justification at the expense of causation thus generating a familiar dilemma between eliminativism and primitivism. As a diagnosis I have argued that the dilemma is predicated on the assumption that reasons are individuated in exclusively causal and normative terms respectively. To avoid this dilemma I have proposed an outline of an alternative approach, ETA that combines the ecological account of agency with the invariance account of teleological explanation.

The central idea is that the theoretical role of reasons in the explanation of an action is to give the ecological role of an action. The ecological role of an action, in turn, is the contribution it makes to the attainment of agent’s goal given her repertoire and affordances. The reason for performing an action is thus the complex ecological fact that performing the action is conducive to the attainment of the agent’s goal. So the reason why the agent caused the action is the fact that acting in that way is a (hypothetically) reliably way to attain the goal and hence is good for attaining the goal. This implies that the agent caused the action for a reason. On account of this ETA avoids the dilemma and offers a complete and naturalistic account of action-explanation that vindicates both its causal and rational character.

The ecological account of action explanation does not give knockout arguments against the standard approaches. Rather, it aims to capture what each of the standard accounts does right while avoiding its problems. On account of this, it must be included in the debate.
Bibliography


